



LIBRARY
New Delhi

Call No. _____

Acc. No. _____

23636

J o u r n a l o f A p p l i e d P h y s i c s

Board of Editors

ELMER HUTCHISSON, *Editor*

Associate Editors

**R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE**

**S. H. CALDWELL
CLARK GOODMAN
SELIG HECHT
W. JAMES LYONS**

**PAUL E. ŠABINE
FREDERICK SEITZ
J. R. TOWNSEND**

VOLUME 17

JANUARY-DECEMBER, 1946

Published by the
AMERICAN INSTITUTE OF PHYSICS
Incorporated

INFORMATION FOR CONTRIBUTORS

Manuscripts

Send manuscripts to Elmer Hutchisson, Case School of Applied Science, Cleveland 6, Ohio.

Papers must be in English, **typewritten double spaced** with wide margins on a durable white paper, preferably lettersize. **The original copy**, not the carbon, should be submitted.

For general style, spellings, abbreviations, and form, recent issues of this journal should be consulted.

Abstracts

An abstract must accompany each article. It should be adequate as an *index* and as a *summary*. As an index it should give all subjects, major and minor, concerning which new information is presented. As a summary it should give the conclusions of the article and all numerical results of general interest.

Mathematical Expressions

Make all expressions clear to the typesetter. Identify in the margin of the manuscript Greek letters and unusual symbols.

Use fractional exponents instead of root signs.

Avoid complicated exponents and subscripts. Should it be necessary to repeat a complicated expression a number of times it should be represented by some convenient symbol.

The solidus (/) should be used wherever possible for fractions.

References

References should appear as footnotes, numbered consecutively, and arranged thus:

¹ A. B. Smith, *Phys. Rev.* **41**, 852 (1932).

² G. R. Harrison, *Atoms in Action* (William Morrow & Company, New York, 1939), first edition, p. 30.

Figures and Tables

A separate sheet listing captions for all figures must accompany manuscript. Each table should have a caption. These in all cases should be complete in themselves so as to make the data intelligible to the reader without reference to the text. Each figure and table should be cited in sequence in the text.

Line Drawings and Photographs

Plan figures for approximately one-half reduction. Wherever possible arrange figures so that after reduction they will not be wider than **three inches**. **Lettering must be large enough to be readable after reduction.**

Line drawings must be made with black India ink on white paper or tracing cloth. Coordinate paper is not desirable, but if used must be blue-lined and all coordinates to be reproduced drawn with India ink.

Drawings which cannot be used will be returned to the author.

Photographs should show clear black and white contrasts.

Avoid round and oval photographs.

Alterations

Take great care that each manuscript is typographically accurate. A limited number of alterations in proof is unavoidable, but the cost of making extensive alterations after the article has been set in type will be charged to the author.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publications Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

INFORMATION FOR SUBSCRIBERS

The *JOURNAL OF APPLIED PHYSICS* is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as *PHYSICS*. It aims to be of service not only in physical laboratories, but also in the laboratories of industrial, chemical, geological, geophysical, meteorological, radio, and similar concerns. The *JOURNAL OF APPLIED PHYSICS* publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Subscriptions and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

SUBSCRIPTION PRICE

	U. S. AND CANADA	ELSEWHERE
To members of scientific societies ..	\$5.00	\$5.70
To all others	7.00	7.70

BACK NUMBER PRICE

Complete set: Vol. 1, July, 1931—Vol. 16, 1945—\$101.50.
Yearly back number rate: \$7.70.
Single copies: \$0.70 each.

Advertising rates will be supplied on request.

American Institute of Physics

57 East 55 Street

New York 22, New York

CONTENTS

JANUARY, 1946

Reconversion of the Journal of Applied Physics	1
Atomic Engineering? THEODORE VON KÁRMÁN	2
Interchange of Energy between an Electron Beam and an Oscillating Electric Field JESS MARCUM	4
On the Improvement of Resolution in Electron Diffraction Cameras JAMES HILLIER AND R. F. BAKER	12
Applications of Metallic Shadow-Casting to Microscopy ROBLEY C. WILLIAMS AND RALPH W. G. WYCKOFF	23
Electrical Resistance of Iron Wires and Permalloy Strips at Radiofrequencies ALVA W. SMITH, J. H. GREGORY, AND J. T. LYNN	33
Liquid Rise in a Capillary Tube WESLEY E. BRITTON	37
Shunt and Series Sections of Transmission Line for Impedance Matching C. T. TAI	44
On the Proportion of Crystalline and Amorphous Components in Stretched Vulcanized Rubber A. J. WILDSCHUT	51
Letters to the Editor:	
On Plane Rotations in n Dimensions PETER T. LANDSBERG	60
Letter to the Editor GERALD PICKELL	61
On a Novel Form of Refrigerator JENNY E. ROSENTHAL	62
Elastic Limit for Dynamic Loading V. BARGMANN AND H. F. LUDLOFF	63
Erratum: Secondary Electron Emission of Pyrex Glass C. W. MUELLER	62
Electron Microscope Society of America. Third Annual Meeting	66
Here and There	64
New Booklets	65
Recent Applications of Physics	xiv

FEBRUARY, 1946

Problems in Non-Elastic Deformation of Metals CLARENCE ZENER AND J. H. HOLLLOMON	69
Problems in Fracture of Metals J. H. HOLLLOMON AND C. ZENER	82
Some Integral Equations of Potential Theory H. BATEMAN	91
Federation of Atomic Scientists	108
Calculation of the Output from Non-Linear Mixers HARRY STOCKMAN	110
Filmless Sample Mounting for the Electron Microscope JOHN H. L. WATSON	121
Investigation of Secondary Phases in Alloys by Electron Diffraction and the Electron Microscope R. D. HEIDENREICH, L. STURKEY, AND H. L. WOODS	127
Errata:	
On the Representation of Rigid Rotations H. SCHWERDIFFERGER	109
On Maximum Gain-Band Width Product in Amplifiers W. W. HANSEN	109
Here and There	103
Calendar of Meetings	109
New Booklets	108

MARCH, 1946

Film-Lubrication between Spherical Surfaces: with an Application to the Theory of the Four-Ball Lubricant Testing Instrument J. HOWLETT	137
Analysis of Multicomponent Mixtures of Hydrocarbons in the Liquid Phase by Means of Infra-Red Absorption Spectroscopy D. L. FRY, R. E. NUSBAUM, AND H. M. RANDALL	150
Wide Range Static Strength Testing Apparatus for Glass Rods T. C. BAKER AND F. W. PRESTON	162
Fatigue of Glass under Static Loads T. C. BAKER AND F. W. PRESTON	170
Effect of Water on the Strength of Glass T. C. BAKER AND F. W. PRESTON	179
Fatigue Modulus of Glass J. L. GLATHART AND F. W. PRESTON	189

Induction Heating of Hollow Metallic Cylinders	ANDREW GEMANT	195
Tensile Properties of Several Special Steels and Certain Other Materials under Pressure	P. W. BRIDGMAN	201
Letter to the Editor:	GEORGE G. BRUCK	213
Here and There		213
New Booklets		214

APRIL, 1946

New Electrolytic Selenium Photo-Cell . A. VON HIPPEL, J. H. SCHULMAN, AND E. S. RITNER	215
Studies of Plastic Flow of Steel, Especially in Two-Dimensional Compression	P. W. BRIDGMAN 225
Method of Determining the Size of Droplets Dispersed in a Gas	R. L. STOKER 243
Accuracy of Lumping in an Electric Circuit Representing Heat Flow in Cylindrical and Spherical Bodies	VICTOR PASCHIS AND MICHAEL P. HEISLER 246
Determination of the Temperature Rise and the Maximum Safe Current through Multiconductor Electric Cables	HAIG P. ISKENDERIAN AND WILLIAM J. HORVATH 255
Computer for Solving Linear Simultaneous Equations	CLIFFORD E. BERRY, DOYLE E. WILCOX, SIBYL M. ROCK, AND H. W. WASHBURN 262
Thin Cylindrical Antenna: A Comparison of Theories	DAVID MIDDLETON AND RONOLD KING 273
Correction of Diffraction Amplitudes for Lorentz and Polarization Factors	M. J. BUERGER AND GILBERT E. KLEIN 285
Further Improvement in the Resolving Power of the Electron Microscope	JAMES HILLIER 307
Preparation of Electron Microscope Specimens for Determination of Particle Size Distribution in Aqueous Suspensions	A. M. CRAVATH, A. E. SMITH, J. R. VINOGRAD, AND J. N. WILSON 309
Calendar of Meetings	284
Here and There	311
New Books	312

MAY, 1946

Editorial	E. H. 313
Elastic Properties of Cork. I. Stress Relaxation of Compressed Cork	S. L. DART AND EUGENE GUTH 314
Formation of Ionized Water Films on Dielectrics under Conditions of High Humidity	ROBERT F. FIELD 318
Thermal Polymerization of Drying Oils	H. E. ADAMS AND P. O. POWERS 325
Electrostatic and Tensile Properties of Rubber and GR-S at Elevated Temperatures	R. S. HAVENHILL, H. C. O'BRIEN, AND J. J. RANKIN 338
Significance of the Equation of State for Rubber	EUGENE GUTH, PAUL E. WACK, AND ROBERT L. ANTHONY 347
Theory of Permanent Set at Elevated Temperatures in Natural and Synthetic Rubber Vulcanizates	R. D. ANDREWS, A. V. TOBOLSKY, AND E. E. HANSON 352
Crystallization of Unvulcanized Rubber at Different Temperatures	LAWRENCE A. WOOD AND NORMAN BEKKEDAHL 362
Tackiness of GR-S and Other Elastomers	W. F. BUSSE, J. M. LAMBERT, AND R. B. VERDERY 376
Applications of Infra-Red Methods in the Structural Examination of Synthetic Rubber	J. E. FIELD, D. E. WOODFORD, AND S. D. GEHMAN 386
Photoelectric Instrument for Light Scattering Measurements and a Differential Refractometer	P. P. DEBYE 392
Thermal Expansion and Second-Order Transition Effects in High Polymers. III. Time Effects	R. S. SPENCER AND R. F. BOYER 398
Abstracts	405
Calendar of Meetings	361
Here and There	408
New Books	408
New Booklets	410

JUNE, 1946

Study of Distortion in Electron Microscope Projection Lenses	JAMES HILLIER	411
Parafocusing Properties of Microcrystalline Powder Layers in X-Ray Diffraction Applied to the Design of X-Ray Goniometers	J. C. M. BRENTANO	420
Study of Oxide Cathodes by X-Ray Diffraction Methods. Part I. Methods, Conversion Studies, and Thermal Expansion Coefficients	A. EISENSTEIN	434
Methods for Betatron or Synchrotron Beam Removal	EUGENE C. CRITTENDEN, JR., AND WILLIAM E. PARKINS	444
Graphical Method for Determining Particle Trajectories	WILLIAM E. PARKINS AND EUGENE C. CRITTENDEN, JR.	447
Analyses of Mixtures of Light Gases by Infra-Red Absorption	NORMAN D. COGGESHALL AND ELEANOR L. SAIER	450
Methods of Minimizing Lead Loss in Emissivity and Resistivity Determinations	L. MALTER AND D. B. LANGMUIR	456
Theory and Application of the Parallel Plate Plastometer	G. J. DIENES AND H. F. KLEMM	458
General Relations of Flow in Solids and Their Application to the Plastic Behavior of Tire Cords	W. JAMES LYONS	472
Fatigue Failure of Rayon Tire Cord	R. C. WALLER AND W. E. ROSEVEARE	482
Dynamic Characteristics of Rubber Supports from Vibration Table Data	ERNEST G. CHILTON	492
Absorption of Microwaves by Gases	W. D. HERSHBERGER	495
Pressure-Volume Relation for Cylindrical Tubes with Elastomeric Walls: The Human Aorta	ALLEN L. KING	501
Approximations to Planckian Distributions	PARRY MOON AND DOMINA EBERLE SPENCER	506
Approximate Calculation of the Range of a Winged Jet Projectile	M. ZHIGMEW KRZYWOBLOCKI	515
Design of Broad Band I.F. Amplifiers	RICHARD F. BAUM	519
Relation between Nodal Positions and Standing Wave Ratio in a Composite Transmission System	EUGENE FEENBERG	530
Erratum: Small Prism Infra-Red Spectroscopy	R. BOWLING BARNES, ROBERT S. McDONALD, AND VAN ZANDT WILLIAMS	532
Calendar of Meetings		514
Here and There		533
New Books		535
New Booklets		540

JULY, 1946

Spark Shadowgraphic Study of Body Waves in Water	J. HOWARD McMILLEN AND E. NEWTON HARVEY	541
Lightning Recorded by Aeroplane Propeller	J. M. BRYANT AND M. NEWMAN	555
Microwave Dielectric Loss Measuring Technique	WILLIAM R. MACLEAN	558
Magnetically-Controlled Gas Discharge Tubes	R. E. B. MAKINSON, J. M. SOMERVILLE, K. RACHEL MAKINSON, AND P. THONEMANN	567
Depreciation of Fluorescent Lamps without Passage of Current	H. C. FROELICH	573
Interpretation of Harker Syntheses	M. J. BUERGER	579
Dissociation Energies of Surface Films of Various Oxides as Determined by Emission Measurements of Oxide Coated Cathodes	HAROLD JACOBS	596
Experimental Investigation of Forced Vibrations in a Mechanical System Having a Non-Linear Restoring Force	CARL A. LUDEKE	603
New Method for Measuring Dielectric Constant and Loss in the Range of Centimeter Waves	S. ROBERTS AND A. VON HIPPEL	610
Interpolation with the Aid of a Plot of First Differences	GORDON S. FULCHER	617
Use of Color and Fluorescence Indicators for Determining the Structure of Glass	W. A. WEYL	628
Calendar of Meetings		639
Here and There		640
New Books		641
New Booklets		641

AUGUST, 1946

Oxide Coated Cathode Literature, 1940-1945	JOHN P. BLEWETT	643
Pulsed Properties of Oxide Cathodes	EDWARD A. COOMES	647
Study of Oxide Cathodes by X-Ray Diffraction Methods. Part II. Oxide Coating Composition	A. EISENSTEIN	654
Studies of the Interface of Oxide Coated Cathodes	A. FINEMAN AND A. EISENSTEIN	663
Midwest Research Institute - A New Scientific Research Institution	GEORGE E. ZIEGLER	669
Method for Preparing Rubber Latex Specimens for the Electron Microscope	R. H. KELSEY AND E. E. HANSON	675
Perturbation Theory of the Normal Modes for an Exponential <i>M</i> -Curve in Non-Standard Propagation of Microwaves	C. L. PEKERIS	678
Stresses in Cylindrical Glass-Metal Seals with Glass Inside	ALBERT W. HULL	685
Particle Size Determination from X-Ray Line Broadening	L. S. BIRKS AND H. FRIEDMAN	687
Effect of Hydrostatic Pressure on Plastic Flow under Shearing Stress	P. W. BRIDGMAN	692
Hysteresis and Related Elastic Properties of Tire Cords	HELMUT WAKEHAM AND EDITH HONOLD	698
Letters to the Editor:		
On the "Power Function"	FIDEL AUSA FUERTES	712
Here and There		673
Calendar of Meetings		674

SEPTEMBER, 1946

Internal Friction and Plastic Extension of Zinc Single Crystals	THOMAS A. READ AND E. P. T. TYNDALL	713
Design of Broad Band I.F. Amplifier. Part II	RICHARD F. BAUM	721
Numerical Solution of Laplace's Equation in Composite Rectangular Areas	MAX M. FROCHT	730
Erratum: Correction of Diffraction Amplitudes for Lorentz and Polarization Factors	M. J. BUEGER AND GILBERT E. KLEIN	743
Thermoluminescence and Conductivity of Phosphors	ROBERT C. HERMAN AND CHARLES F. MEYER	743
Oscillation Conditions in Single Tuned Amplifiers	WILLIAM R. FAUST AND HUGO M. BECK	749
Letter to the Editor:		
Evaporation of Antimony	LOUIS HARRIS	757
Calendar of Meetings		756
Here and There		757

OCTOBER, 1946

Supplementary Bibliography of Electron Microscopy	MARY E. RATHBUN, MARY J. EASTWOOD, AND ORLAN M. ARNOLD	759
Electronic Spectroscopy	GEORGE C. SZIKLAI AND ALFRED C. SCHROEDER	763
Calculated Frequency Spectrum of the Shot Noise from a Photo-Multiplier Tube	R. D. SARD	768
Response of Biased, Saturated Linear, and Quadratic Rectifiers to Random Noise	DAVID MIDDLETON	778
Flow of Gases in Pipes at Low Pressures	GORDON P. BROWN, ALBERT DiNARDO, GEORGE K. CHENG, AND THOMAS K. SHERWOOD	802
Absorption of Microwaves by Gases. II	JOHN E. WALTER AND W. D. HERSHBERGER	814
Electrical Observations of the Austenite-Martensite Transformation in Steel	ANDREW W. McREYNOLDS	823
Energy Build-Up in Magnetrons	LLOYD P. HUNTER	833
Measured Impedance of Cylindrical Dipoles	D. D. KING	844
Elementary Treatment of Longitudinal Debunching in a Velocity Modulation System	EUGENE FEENBERG	852
Calendar of Meetings		856
Here and There		856

NOVEMBER, 1946

Theoretical Physics in the Metallurgical Laboratory of Chicago	E. P. WIGNER	857
Composition and Diffraction Effects in X-Ray Microradiographs	R. SMOLUCHOWSKI, C. M. LUCHT, AND J. M. HURD	864
Transient Analysis of Linear Systems, Using Underwater Explosion Waves	M. F. M. OSBORNE AND J. L. CARTER	871
X-Ray Method for Measuring the Thickness of Thin Crystalline Films	A. EISENSTEIN	874
New Bridge Photo-Cell Employing a Photo-Conductive Effect in Silicon. Some Properties of High Purity Silicon	G. K. TEAL, J. R. FISHER, AND A. W. TREPTOW	879
Production of High Centrifugal Fields	J. W. BEAMS, J. L. YOUNG, III, AND J. W. MOORE	886
Cathode-Ray Tube for Viewing Continuous Patterns	J. B. JOHNSON	891
Improvements in the Stability of the FP-54 Electrometer Tube	J. M. LAFFERTY AND K. H. KINGDON	894
Picturesque Oxidation of Molybdenum Vacuum-Tube Leads	HERMAN A. LIEBHAFSKY	901
Luminescence of $(\text{Zn, Be})_2\text{SiO}_4$: Mn and Other Manganese-Activated Phosphors	JAMES H. SCHULMAN	902
Recent Ozone Investigations	ARTHUR W. EWELL	908
Properties of Welded Contact Germanium Rectifiers	H. Q. NORTH	912
Fine Structure of Viscose Rayon	H. G. INGERSOLL	924
Theoretical Comparison of the Visual, Aural, and Meter Reception of Pulsed Signal in the Presence of Noise	J. H. VAN VLECK AND DAVID MIDDLETON	940
Permeability of Different Rubbers to Gases and Its Relation to Diffusivity and Solubility	G. J. VAN AMERONGEN	972
Calendar of Meetings		900
Here and There		986
New Booklets		988

DECEMBER, 1946

Report of the Electron Microscope Society of America's Committee on Resolution		989
Electron Microscope Examination of the Microphysical Properties of the Polymer Cuprene	JOHN H. L. WATSON AND K. KAUFMANN	996
Machine for the Application of Sand in Making Fourier Projections of Crystal Structures	DAN McLACHLAN, JR. AND E. F. CHAMPAGNE	1006
Radar Reflections from Long Conductors	F. BLOCH, M. HAMERMESH, AND M. PHILLIPS	1015
Viscometric Investigation of Dimethylsiloxane Polymers	ARTHUR J. BARRY	1020
Theory of Small Signal Bunching in a Parallel Electron Beam of Rectangular Cross Section	EUGENE FEENBERG AND DAVID FELDMAN	1025
On Vibrations of Shallow Spherical Shells	ERIC REISSNER	1038
Non-Uniform Transmission Lines and Reflection Coefficients	L. R. WALKER AND N. WAX	1043
Resonant Frequencies of the Nosed-In Cavity	ERNEST MAYER	1046
Study of a Certain Type of Resonant Cavity and Its Application to a Charged Particle Accelerator	E. S. AKLEY	1056
Millimeter-Wave Reflex Oscillator	J. M. LAFFERTY	1061
Resonant Circuit Modulator for Broad Band Acoustic Measurements	GORDON FERRIE HULL, JR.	1066
Transient Temperatures around Heating Pipes Maintained at Constant Temperature	ANDREW GEMANT	1076
Effect of Temperature on the Strength and Fatigue of Glass Rods	B. VONNEGUT AND J. L. GLATHART	1082
Plastic Flow, Creep, and Stress Relaxation. Part I. Plastic Flow	CHARLES MACK	1086
Plastic Flow, Creep, and Stress Relaxation. Part II. Creep	CHARLES MACK	1093
Plastic Flow, Creep, and Stress Relaxation. Part III. Creep and Elastic After-Effect	CHARLES MACK	1101
Asymptotic Solutions for the Normal Modes in the Theory of Microwave Propagation	C. L. PEKERIS	1108

Letters to the Editor:

Concerning "Computer for Solving Linear Simultaneous Equations"	
. C. C. EAGLESFIELD; CLIFFORD E. BERRY	1125
Preparing Pigments for the Electron Microscope	A. M. CRAVATH 1125
Coordination Number in Solid Solutions - A Mathematical Paradox	DOUGLAS RENNIE HUDSON 1126
Note on a Paper of H. Bateman	F. W. J. OLVER 1127
Author Index to Volume 17	1132
Subject Index to Volume 17	1136
Volume Table of Contents	III
Calendar of Meetings	1005
Here and There	1128
New Books	1130
New Booklets	1131
Index to Advertisers	xxxii

Journal of Applied Physics

Volume 17, Number 1

January, 1946

Reconversion of the *Journal of Applied Physics*

GRADUALLY the *Journal of Applied Physics* is recovering some of its prewar characteristics. During the past few years paper restrictions were particularly acute because of the large increase in circulation which occurred after the paper quota for the *Journal* was set. In fact, had it not been for the generous cooperation of the editors of *The Review of Scientific Instruments* and the *Journal of Chemical Physics* in sharing their paper quotas, several numbers of this *Journal* could not have appeared. However, the size of the past few issues shows that that difficulty is now over, and in this issue we are permitted once more to return to a high quality coated paper stock.

New problems are on the horizon. The number of papers submitted is large and shows signs of increasing. These papers are generally of high quality and cover a wide range as far as subject matter is concerned. If the number of papers increases at a faster rate than the circulation, the Editor may be forced to choose between outstanding papers in Field A as opposed to outstanding papers in Field B. Or he may find it necessary to accept good research contributions at the expense of review or summary articles. Indications of reader preference on these matters would be as welcome as they are scarce.

Many of the engineering societies are in the process of examining their reader interest rather closely to assist in setting definite limits to the fields which they are to cover and to reduce overlapping with other societies. These studies are of considerable interest to this Editor because of the close relation between applied physics and engineering. It is helpful, in visualizing the publishing activities in America of the entire field of technical and engineering knowledge, to make use of a polar chart in which basic physics is at the center. The major subdivisions

of physics, such as mechanics, thermodynamics, acoustics, optics, electrostatics, electrodynamics, may be thought of as occupying particular sectors. Thus, reading out radially in any one sector, for example that devoted to mechanics, one would start with fundamentals in mechanics, contributions to which are usually published in *The Physical Review*. Then one would pass through a region covered by the *Journal of Applied Physics* or the *Journal of Applied Mechanics*. Out still farther one would find subject matter covered by *Mechanical Engineering*, the official publication of the A.S.M.E. The polar chart has the advantage that the sector widens out as the variety of applications increases. In the outer regions we would find mechanics divided into many subjects, such as machinery, automotive engineering, aeronautical engineering, and so forth. Still farther out, perhaps, would be fields covered by trade publications.

In a chart of the type here visualized, the region to be covered by the *Journal of Applied Physics* seems rather naturally to be an annulus surrounding basic physics. In some sectors this annulus becomes very narrow, and in others it must be quite broad, to cover the gap between fundamental physics and engineering. It seems quite clear that one important function of the *Journal* is to provide a publishing medium in those sectors where this gap is broad. Gradually as applications develop, specialized journals will necessarily begin to take over. However, as new fundamental knowledge develops, new sectors will open for the *Journal of Applied Physics*. Thus, the emphasis in the *Journal* is very likely to change as time goes on, but it will always be just outside that covered by those journals concerned with developments in fundamental or "pure" physics.

Atomic Engineering?*

THEODORE VON KÁRMÁN

California Institute of Technology, Pasadena, California

IT has happened now for the third time in our lifetime ("our" referring to the elders of the engineering profession) that new avenues of engineering activities of tremendous importance have been opened by physicists, and the engineering profession has found itself unprepared to challenge the task.

In 1887 the physicist Heinrich Hertz discovered "wireless telegraphy." In fact, he experimentally confirmed the existence of electromagnetic waves which anybody may easily read in Maxwell's equations of the electromagnetic field, i.e., anybody who can read equations at all. It took several decades for radio engineering to become a branch of the engineering profession and for adequate measures to be taken to incorporate the fundamentals of field theory into the engineering curriculum.

The discovery of electromagnetic waves of long wave-length was followed very soon by the discovery of x-rays, i.e., electromagnetic waves of very short length of a frequency far beyond the spectrum of visible light. It has to be said to the credit of the engineering profession that it took advantage of x-ray techniques in various industrial fields, and the design and construction of x-ray apparatus were recognized as an important branch of engineering.

Similarly, the infra-red technique, i.e., the technique of waves longer than those of visible light but very much shorter than radio waves, was justly considered as an engineering problem.

However, in the period between the two world wars, physicists made great efforts to narrow the gap between radio and infra-red waves. They succeeded in producing microwaves which made possible the marvelous technique now commonly designated as the art of radar. The physical facts were known a long time before the beginning of the recent war. However, during the war, physicists, by their superior knowledge of electronics, i.e., the mechanics of electric particles, were able to solve the engineering problems

of producing and directing microwaves. I will not contest the fact that a great many American engineers were sufficiently familiar with electronics; many made important discoveries and contributions to electronic devices. But somehow our engineering education does not encourage the idea of the unlimited horizon, when fundamental thinking in novel fields is required.

And now it seems we are at the threshold of the new atomic age. I do not know whether or not this is true, but certainly we shall have "atomic engineering" in the fields of power and transportation. Are we prepared for the problems involved?

The first application of atomic engineering was directed by military engineers; plants and many gadgets were probably designed and constructed by engineers. However, not only the discovery of the fundamental facts but also the methods of application were suggested, worked out, and tested by physicists. I concede that security regulations and personal relations in the "nuclear clan" played a great part in the choice of collaborators. However, I raise the question, do we give today to the future engineer enough fundamental knowledge in basic questions of the structure of matter, the nature of energy, the relation between matter and energy, so that he will be able to think intelligently in the new field, to have good engineering judgment on possibilities and methods? Is he better equipped than anybody else who reads a few popular articles in the *New York Times* or in the *Saturday Evening Post*? Is there something wrong in our engineering education? I am afraid there is.

First, I believe we have a tendency to restrict our teaching to scientific knowledge which has immediate applications. We often forget that almost every new physical or chemical discovery *might* have engineering application. After all, engineering is the conquest of nature, i.e., matter and energy, for human comfort, and therefore an engineer cannot know too much about the inner structure of the matter against whose whimsical moods he struggles and the laws

* Reprinted by permission from *Mechanical Engineering*, October, 1945.

of energy which he wants to exploit and harness.

Second, we underestimate the interest of our students in "natural philosophy." We are reluctant to present the fundamentals of physics and chemistry as a living science full of question marks and changing concepts. We believe we ought to introduce the findings of research only after the unshakable truth has been established. We tend to stick to classical concepts. I found, while teaching mechanics of continuous media, elasticity, and fluid dynamics, that my students listened eagerly as I told them something about atomic structure of the materials and kinetic theory of gases. I wonder how many engineering students obtain a picture of entropy, chemical reaction, and the like, from a modern point of view. And why should ordinary combustion be an engineering topic, and nuclear reaction a mystery of modern alchemy?

Third, we overestimate the importance of transfer of experience to coming engineering generations in comparison with a training aimed at true understanding of the happenings of nature. We try to train engineers so that the employer can use them almost immediately after graduation from school. One of my former students was employed as an instructor in a well-known engineering school. He suggested some changes in the hydraulics curriculum in the direction of modern fluid dynamics. His superior asked him the use of teaching the findings of research men who, he conceded, found interesting results in hydrodynamics. None of these men, he said, would be able to design a sewage pipe system. No doubt he was right in this statement of fact, but he was wrong in his conclusion. Fortunately, the more intelligent employers are beginning to realize that immediate usefulness is not the most important criterion of a novice engineer. They appreciate sound fundamental understanding and do not want the school to train for them the type of practical engineer, who—according to the *bon mot* of a prominent Englishman, himself an engineer—perpetuates the errors of his predecessors.

Fourth, I certainly do not wish to make of every engineer a scientist, or, God forbid, a nuclear physicist. However, we shall try to give him an education which shall enable him to follow the progress of science, as, I believe,

every medical doctor should have sufficient education to follow the development of biology which has produced in our time almost as spectacular discoveries as physical science. At the institution at which I have been teaching in the last fifteen years, science and engineering students have essentially common curricula the first two years. The physics student continues after that to move on a relatively high intellectual level of understanding natural phenomena from a scientific point of view. I have often wondered why the engineering student has to make a steep dive into an atmosphere in which the beam on three supports is considered a most difficult problem and the exact clear principles of mechanics and thermodynamics are replaced by semi-scientific, semi-popular "ersatz" truths.

I am convinced that the knowledge of the deepest origins and also the limitations of the principles do not handicap a person in their practical application; as a matter of fact, real knowledge makes the application easier and safer. And after all, physics students are not any more intelligent than engineering students.

I have digressed from the subject indicated by the title. From the little I know and the little I have read on the subject, I really believe that the problems involved in the development of atomic power for stationary and transportation purposes are primarily of the nature of engineering and especially applied mechanics. In other words, the picture of the nuclear processes obtained by the physicists contains so much of the truth, if such thing as truth exists, that from now on systematic observation and computation must lead to practical solutions. Certainly intriguing engineering problems, like the development of materials resistant to extreme temperatures and extreme corrosion, and difficult applications of our knowledge in diffusion and heat transfer, will be involved. However, it may sound paradoxical, but it seems to me that the processes involved in atomic engineering are less complex than, for example, the processes of combustion of conventional solid and liquid fuels; less complex in the sense that simple considerations and theoretical computations may give closer approximations to reality than in the case of molecular reactions. What can appeal more to a scientific engineer?

Contributed Original Research

Interchange of Energy between an Electron Beam and an Oscillating Electric Field

JESS MARCUM*

Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania

(Received October 25, 1945)

Relations between various parameters are obtained which describe the behavior of an accelerated electron beam which is caused to traverse an alternating electric field. In particular, a mechanicographical means for obtaining the gain or loss of energy is described. It is shown that under the most favorable conditions a maximum of 17 percent of the energy in the accelerated beam may be transferred to the alternating field. Application of these principles to a type of ultra-high frequency oscillator is treated.

INTRODUCTION

THIS report deals with a theoretical analysis of the conversion of kinetic energy of electrons into radio frequency energy of an oscillating electric field. In Fig. 1, *a* is a cathode maintained at a negative potential V_0 with respect to the plates *b* and *c*. In the region between *b* and *c* there is an oscillating electric field so the electric vector is uniform and equal to $E \sin \omega t$, being directed along the direction of motion of the electrons.

In the theory developed in this paper space charge is neglected, each electron is supposed to go all the way through without striking the grids, and non-uniformity of the electric field near the grids is neglected. In practice, the plates *b* and *c* may be built into a cavity resonator but this is not essential for the first part of the discussion which follows.

The problem of the energy exchange between an electron beam and an oscillating field disposed in this way relative to the beam has been treated analytically by Müller and Rostas.¹ But the analytical method which they use involves approximations which limit the range of results they can obtain.

In this report it is shown how the problem can be studied by means of a mechanical-graphical

computing device which gives results of more generality than those found by Müller and Rostas.

METHOD OF SOLUTION

The equation of motion of an electron entering at phase δ is

$$\frac{d^2x}{dt^2} = \frac{eE}{m} \sin(\omega t + \delta), \quad (1)$$

where e is electronic charge, m is electronic mass, and E is the amplitude of the electric vector. The electrons all enter with velocity v_0 determined by the accelerating cathode potential V_0 . Integrating (1) we find for the velocity v at later times

$$v = v_0 - \frac{eE}{m\omega} [\cos(\omega t + \delta) - \cos \delta]. \quad (2)$$

A second integration gives the coordinate x as a

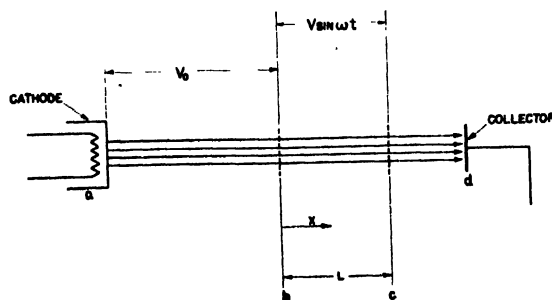


FIG. 1. Schematic diagram.

* On leave of absence from Westinghouse Research Laboratories. Present address, Princeton University, Princeton, N. J.

¹ J. J. Müller and E. Rostas, "Un Générateur à Temps de Transit, Utilisant un Seul Résonateur de Volume," *Helv. Phys. Acta* 13, 435 (1940).

function of the time,

$$x = v_0 t - \frac{eE}{m\omega^2} [\sin(\omega t + \delta) - \omega t \cos \delta - \sin \delta]. \quad (3)$$

It is convenient to introduce certain dimensionless variables as follows,

$$\eta = \frac{eE}{m\omega v_0} = \text{ratio of momentum change produced by force } eE \text{ in time } 1/\omega \text{ to the initial momentum of the electron.}$$

$$\theta = \omega t + \delta = \text{phase of oscillating field in "electrical radians."}$$

$$\alpha = \frac{\omega x}{2\pi v_0} = \text{time measured in cycles of oscillation frequency to go the distance } x \text{ when moving with the entrance velocity, } v_0.$$

In terms of these quantities (2) and (3) become

$$v/v_0 = 1 - \eta [\cos \theta - \cos \delta], \quad (2b)$$

$$2\pi\alpha = \theta - \delta - \eta [\sin \theta - \sin \delta + (\delta - \theta) \cos \delta]. \quad (3b)$$

If we could eliminate θ between these two equations, we could express (v/v_0) as a function of α , which is a measure of the distance gone in the oscillating field, and δ the phase of entry into the field, as well as η which is a measure of the amplitude of the oscillating field.

The ratio of the decrease in kinetic energy of an electron to its original energy is

$$\rho(\delta) = \frac{mv_0^2/2 - mv^2/2}{mv_0^2/2} = 1 - \left(\frac{v}{v_0}\right)^2, \quad (4)$$

which, since v/v_0 is a function of η , δ , and α , becomes also a function of the same variables which will be designated

$$\rho(\delta, \eta, \alpha). \quad (5)$$

If the beam entering the grid at b is unmodulated, then equal numbers of electrons enter at all phases of the oscillating field and the average energy exchange of the electrons with the field is given by averaging $\rho(\delta, \eta, \alpha)$ over all phases δ in a cycle.

We denote this average, which depends parametrically on η and α by $\bar{\rho}(\eta, \alpha)$ in accord with the notation,

$$\bar{\rho}(\eta, \alpha) = \frac{1}{2\pi} \int_0^{2\pi} \rho(\delta, \eta, \alpha) d\delta. \quad (6)$$

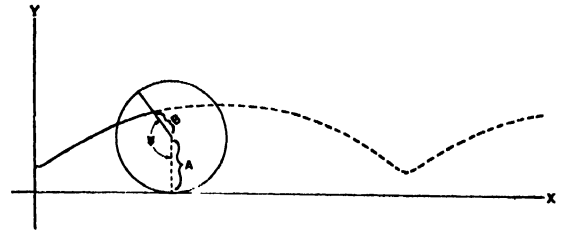


FIG. 2. Generation of a cycloid.

A positive value of $\bar{\rho}(\eta, \alpha)$ means that on the average the electrons lose energy to the oscillating field, and $\bar{\rho}(\eta, \alpha)$ gives the fraction of the beam energy that is given up to the electric field. In applying the formula for $\bar{\rho}(\eta, \alpha)$, the quantity α is to be calculated by using for x the distance of separation of the two grids b and c .

In actual fact, however, there is no exact analytical way to eliminate θ between the Eqs. (2b) and (3b) in order to carry out the calculation as formulated. Müller and Rostas give an approximate analytical calculation of the problems.

Equations (2b) and (3b) can be written in this form

$$v/v_0 = (1 + \eta \cos \delta) - \eta \cos \theta, \quad (2c)$$

$$2\pi\alpha = (1 + \eta \cos \delta)\theta - \eta \sin \theta - [\delta(1 + \eta \cos \delta) - \eta \sin \delta]. \quad (3c)$$

These equations are evidently the equations representing cycloids in a graph in which (v/v_0) is plotted as ordinate and $2\pi\alpha$ is plotted as abscissa. The standard equations for the cycloid generated by rolling a circle of radius A on the x axis with the tracing point at radius B from the center of the circle are,

$$\begin{aligned} X &= A\psi - B \sin \psi \\ Y &= A - B \cos \psi \end{aligned} \quad (7)$$

Comparing with (2c) and (3c) we see that the relationship between (v/v_0) and $2\pi\alpha$ is given by a cycloid in which

$$\begin{aligned} (1 + \eta \cos \theta) &= \text{radius of rolling circle.} \\ \eta &= \text{distance of tracing point from center.} \\ \theta &= \text{angle through which rolling circle is turned.} \end{aligned}$$

$$[\delta(1 + \eta \cos \theta) - \eta \sin \delta] = \text{displacement of origin along } x \text{ axis.}$$

The relations are indicated in Fig. 2.

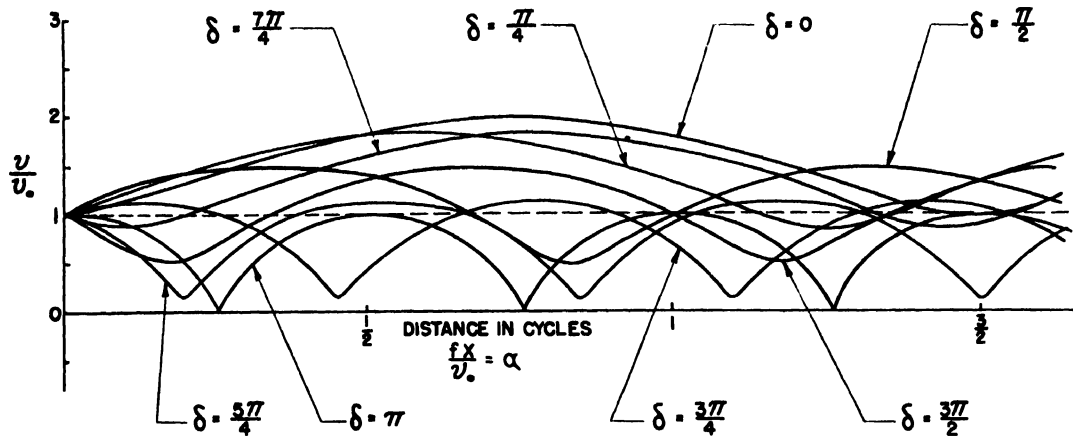


FIG. 3. Cycloidal graph for $\eta = 0.5$.

Because of this fact, a convenient means is available for getting at the velocity of the electrons as a function of distance gone in the oscillating field. It is easy to cut out wheels of the appropriate radii for a given value of η and different phases of entry δ and to roll out the corresponding cycloids. This is what was actually done to obtain the results presented in this paper. In practice, eight equally spaced values of δ throughout a cycle were used. The values of $\bar{\rho}$ were then calculated by averaging the energy exchange for these eight phases as an approximation to integration over all phases.

A section of a typical family of cycloids is shown in Fig. 3. These correspond to the eight different phases of entry in the special case $\eta = \frac{1}{2}$. It may be shown that we must have $\eta < 0.823$ in order that none of the electrons will be returned into the accelerating field. (See Appendix.) We do not consider any cases for which $\eta > 0.823$ in this paper. The calculations for this condition are considerably more involved.

SIMPLIFIED QUALITATIVE PICTURE

Since the main object of the study is to learn the conditions of operation under which $\bar{\rho}(\eta, \alpha)$ is as great as possible, it is desirable to give a rough consideration of the way in which the conversion efficiency depends on these two parameters. In order to make $\bar{\rho}$ as great as possible it would be desirable to have each electron leave the oscillating field when its velocity is least.

For a given value of η , the value of $\langle v/v_0 \rangle$ in (2c) is clearly least when $\cos \theta = 1$, so the least

value of v/v_0 is

$$\left(\frac{v}{v_0} \right)_{\min} = 1 - \eta(1 - \cos \delta). \quad (8)$$

If every electron could be made to leave the oscillating field with this speed ratio (which is not possible in fact), then we should have

$$\bar{\rho} = 1 - \frac{1}{2\pi} \int_0^{2\pi} [1 - \eta(1 - \cos \delta)]^2 d\delta = 2\eta - 3\eta^2/2, \quad (9)$$

which is therefore an absolute upper limit to the conversion efficiency. In Fig. 4 this upper limit is

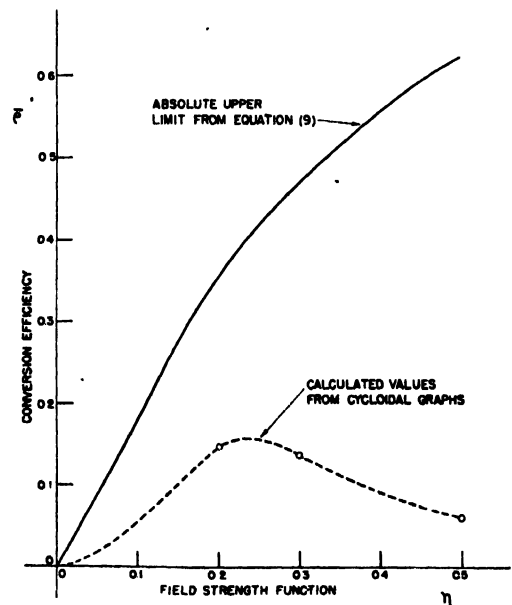


FIG. 4. Efficiency functions.

graphed as a function of η , together with the actual maximum conversion efficiencies (for best possible choice of α at each value of η) obtained from the cycloidal graphs.

We wish to consider further the most favorable values of η and α . The electrons entering at phase $\delta=0$ never have a velocity less than v_0 so they have a conversion efficiency ratio of 0.00 at best. In order to take them out of the field at a time when they have not actually gained energy from the oscillating field the distance x must be chosen to have the values $x=l$ where

$$\omega l/v_0 = 2\pi n_1(1+\eta), \quad (10)$$

where n_1 is any integer. The electrons which enter at phase $\delta=\pi$ may, for appropriate choice of $\omega l/v_0$, be made the most efficient at giving up energy to the field. From (2c) and (3c) it is seen that this appropriate choice for them is

$$\omega l/v_0 = 2\pi(n_2 + \frac{1}{2})(1-\eta), \quad (11)$$

where n_2 is any integer.

We can get a rough formula for the best values of η by requiring (10) and (11) to be consistent, that is, that l be such that the "best" ($\delta=\pi$) and the "worst" ($\delta=0$) electrons leave when their energies are minimum. This gives

$$2\pi n_1(1+\eta) = 2\pi(n_2 + \frac{1}{2})(1-\eta),$$

or

$$\eta = \frac{2(n_2 - n_1) + 1}{2(n_2 + n_1) + 1}. \quad (12)$$

Although this is a very rough calculation,* the

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8
1	$\eta = .300$ $\alpha = 1.80$							
2	.428 .143	.111 .222						
3		.272 .234	.077 .323					
4		.384 .277	.200 .360	.056 .424				
5		.466 .253	.294 .404	.166 .324	.046 .324			
6			.363 .410	.236 .496	.130 .565	.046 .624		
7			.428 .429	.304 .522	.200 .600	.111 .647	.056 .729	
8				.476 .544	.340 .600	.229 .696	.178 .768	.031 .829

FIG. 5. Approximate points for maximum conversion efficiency.

* It is possible to make similar calculations for a two-cavity klystron which give results in good agreement with the more exact theory.

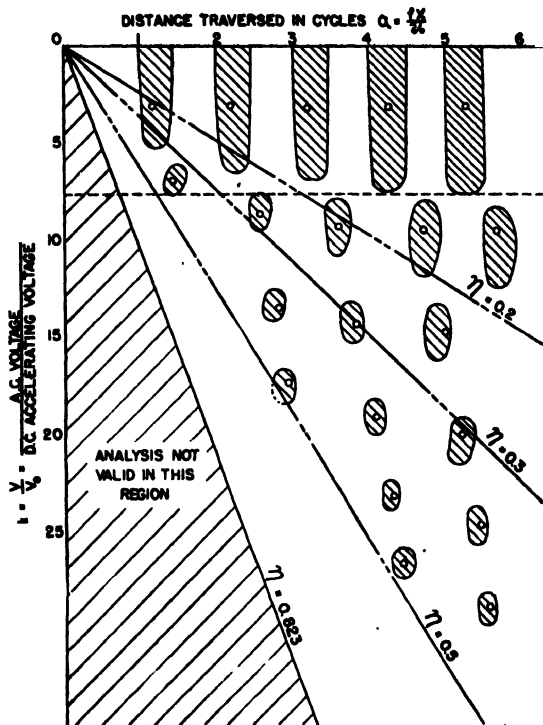


FIG. 6. Regions of positive and negative conversion efficiency.

more exact work with the cycloids bears out the expectation that this consideration should indicate the most favorable values of η . Going with each such value of η will be an appropriate value of α given by (10) or (11).

Each pair of values of n_1 and n_2 will be called a "mode" of operation. In Fig. 5 is a tabulation of the values of η (first line in box) and of α going with quite a range of modes leading to values of η in the range $0 < \eta < \frac{1}{2}$.

Instead of characterizing the operating points by η and α it is convenient to use $k = 4\pi\alpha\eta$ and treat the different operating conditions as defined by the value of k and α . It will be recalled that α is the number of cycles of oscillation required by an electron traveling with the entrance velocity to traverse the space between the grids. We have

$$k = 4\pi\alpha\eta = \frac{eEl}{mv_0^2/2},$$

so the physical significance of k is that it is the ratio of the energy which an electron would gain in the field, if it were static, to its initial energy; k is also the ratio of the oscillating voltage

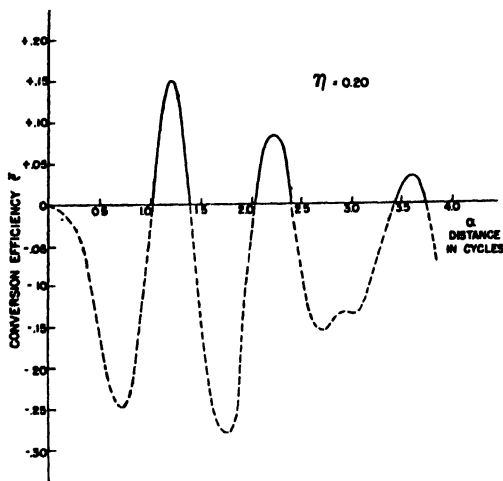


FIG. 7. Conversion efficiency as obtained from cycloidal graphs for $\eta = 0.20$.

amplitude $V = El$ to the initial cathode voltage $V_0 = mv_0^2/2e$.

In Fig. 6 the field of operating conditions is shown in a plot of k against α . The points corresponding to different modes listed in Fig. 5 are shown as "bull's-eyes." In the neighborhood of each bull's-eye is a shaded region. These indicate the fields of values within which $\bar{\rho}$ is positive, that is, in which the beam gives up energy to the oscillating field. The word "mode" is conveniently associated with a whole shaded region surrounding each bull's-eye.

In the regions that are not shaded in Fig. 6 the conversion ratio $\bar{\rho}$ is negative which means that on the average the electrons leave the oscillating field with more energy than they had on entering.

RESULTS OF CYCLOID GRAPHS

Actual values of $\bar{\rho}$ could be shown on an extended form of Fig. 6 in which the values of $\bar{\rho}$ are indicated by means of a contour map. The values of $\bar{\rho}$ are obtained from the cycloids for a number of values of η . The value of $\bar{\rho}$ is first computed for a number of successive values of α at constant η . On Fig. 6 this corresponds to a cross section along one of the diagonal lines as shown. Such a cross section of the $\bar{\rho}$ as a function of k and α surface is shown in Fig. 7. This shows $\bar{\rho}$ as a function of α for $\eta = 0.2$ as calculated from the cycloids.

The necessary wooden wheels were made and rolled and values calculated to construct similar curves for $\eta = 0.3, 0.5$, and 0.823 and the resulting

curves for $\bar{\rho}$ as a function of α are shown in Figs. 8-10.

In Fig. 11 is shown a contour map indicating the dependence of $\bar{\rho}$ on k and α over a region corresponding to the upper left-hand corner of the larger region indicated in Fig. 6. In drawing this graph the data obtained by rolling out the cycloids were supplemented by calculations based on the analytical work of Müller and Rostas. It must be admitted that the exact extent of the region of positive $\bar{\rho}$ surrounding the bull's-eyes going with larger values of k is largely conjectural. By drawing the cycloids for more closely spaced values of η one could find these boundaries more accurately if any real interest attached to a knowledge of them.

It is interesting to consider qualitatively how it comes about that $\bar{\rho}$ is negative over so much more of the (k, α) plane in Fig. 6 than the part in which $\bar{\rho}$ is positive. This comes about partly because more than half of a cycloid lies above its middle. For example, an electron which enters at $\delta = \pi/2$ will have $v > v_0$ for more than half of the distance traversed on a long path even though its velocity varies sinusoidally with the time. A more important factor lies in the fact that the relative energy varies as $(v/v_0)^2$ and so parts of the curve above $v/v_0 = 1$ contribute more than parts below

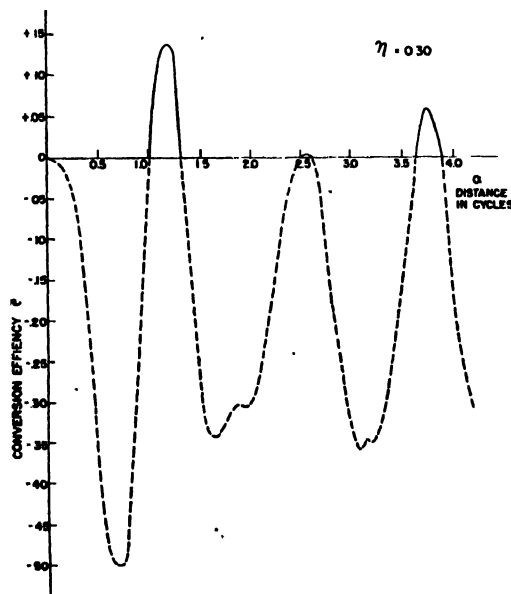


FIG. 8. Conversion efficiency as obtained from cycloidal graphs for $\eta = 0.3$.

by the same amount toward giving a negative value of $\bar{\rho}$.

APPLICATION TO A RESONANT CAVITY OSCILLATOR

Now that the $\alpha, k, \bar{\rho}$ contour map has been obtained (at least over the limited range covered by Fig. 11), one will naturally ask to what practical use can it be put? We limit our discussion to one case, that of constructing an ultra-high frequency oscillator for use in the centimeter wave region. In this case, the planes b and c of Fig. 1 become part of the two opposite walls of a cavity resonator. Any resonator has a property of its dimensions and frequency termed the shunt resistance, S . This is such a quantity that $V^2/2S$, where V is the peak alternating voltage between two specified points of the resonator (usually the points at which the electron beam enters and leaves), gives the average power loss in the walls.

The problem is: Given a tube with a cavity of known resonant frequency f , shunt resistance S , beam current I_0 , beam voltage V_0 , where on the $\alpha, k, \bar{\rho}$ map is the operating point? Provided no power is being taken from the resonator, the following is the condition for equilibrium

$$V_0 I_0 \bar{\rho} = V^2/2S \quad (13)$$

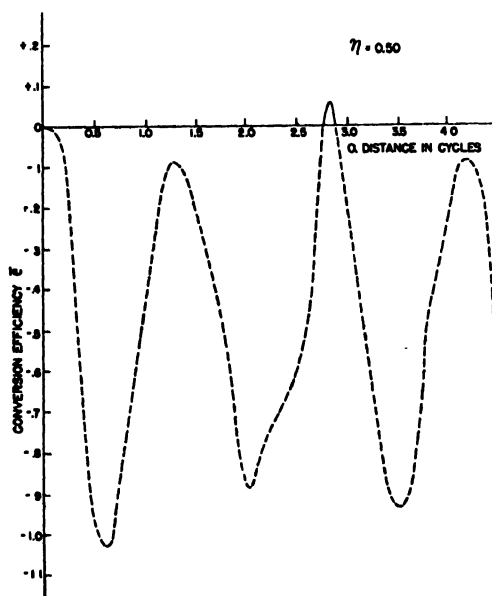


FIG. 9. Conversion efficiency as obtained from cycloidal graphs for $\eta = 0.5$.

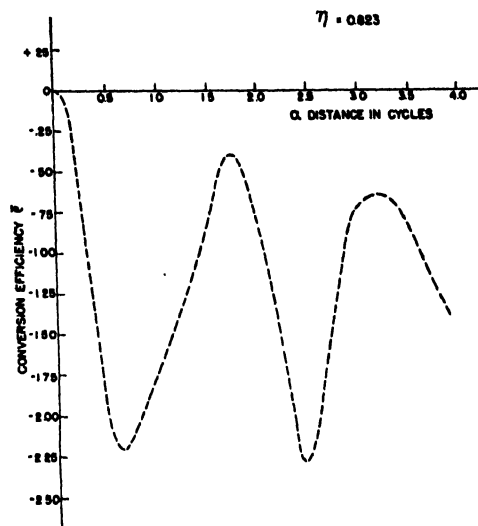


FIG. 10. Conversion efficiency as obtained from cycloidal graphs for $\eta = 0.823$.

which may be transformed to

$$\bar{\rho} = \frac{k^2}{2} \left(\frac{R_0}{S} \right), \quad (14)$$

where $R_0 = V_0/I_0$ is the "resistance" of the beam.

We may now draw a set of curves on the $\alpha, k, \bar{\rho}$ contour map for various values of S/R_0 (the dotted curves on Fig. 11).

Suppose now a tube is in operation; the following quantities are known: f, V_0, I_0 , and S . Compute S/R_0 and α . Drop down from the proper value of α on Fig. 11 to the appropriate S/R_0 curve. This is the operating point, and at once $\eta, \bar{\rho}$, and V are known. Thus, for any set of parameters, the operating conditions may be determined by use of Fig. 11 (or an expanded version).

A tacit assumption in the above reasoning should be pointed out. The frequency at which the tube oscillates has been assumed to be the same as the resonant frequency of the cavity. Actually, the beam current will contain a reactive component which is out of phase with the cavity voltage. This will cause the operating frequency of the tube to shift slightly so that the shunt impedance also contains a reactive component. The error in the efficiency as given in Fig. 11 due to this effect may be shown to be very small. The only point of practical interest attaching to this effect is that for every point on the $\alpha, k, \bar{\rho}$ contour

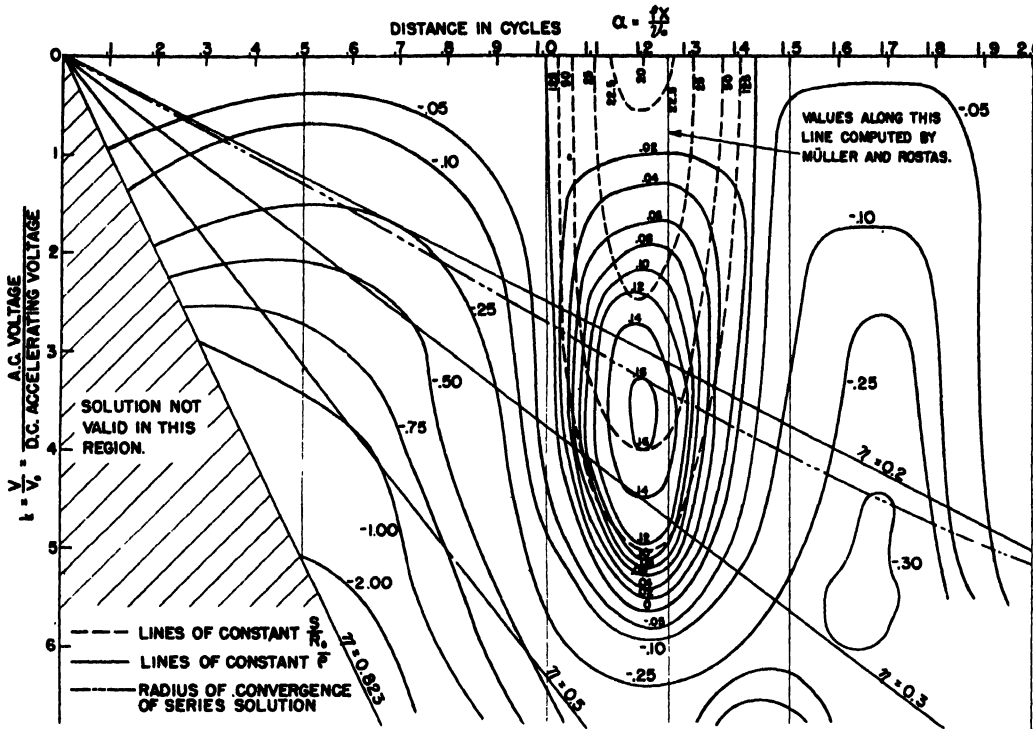


FIG. 11. Complete operating characteristics for first mode and surroundings.

map there is a definite frequency of oscillation. In other words, a second set of frequency contours should be superimposed on Fig. 11 to give complete operating information. The calculation of these frequency contours is possible, but is not attempted here.

We note that S/R_0 must be at least 20, in the first mode, before oscillations are possible. This is in accord with a general principle of operation of velocity-modulated oscillators— for a given tube with a given accelerating voltage, oscillations will start only after the beam current reaches a finite value which is called the “starting current.”

The value of R_0/S for starting is shown (see Appendix II) to be

$$\frac{R_0}{S} = \frac{1}{2} \left(\frac{\sin 2\pi\alpha}{2\pi\alpha} - \frac{\sin^2 \pi\alpha}{\pi^2 \alpha^2} \right). \quad (15)$$

This gives the necessary value of S/R_0 for any α along the $k=0$ axis in order that oscillations will be possible. This information is shown in graphical form in Fig. 12. In effect, it gives the values of the S/R_0 curves of the $\alpha, k, \bar{\rho}$ contour map along the $k=0$ axis.

After the necessary value of starting current is

reached, oscillations will then build up in the first row of modes (Fig. 11) as the current is further increased. However, none but the first row of modes will be self-starting.

If one now desires to take out of the resonator a certain average power, P , it is only necessary to increase the beam current by an amount such that

$$P = V_0 I_0 \bar{\rho} - k^2 V_0^2 / 2S,$$

where $\bar{\rho}$ is the conversion efficiency as determined from Fig. 11.

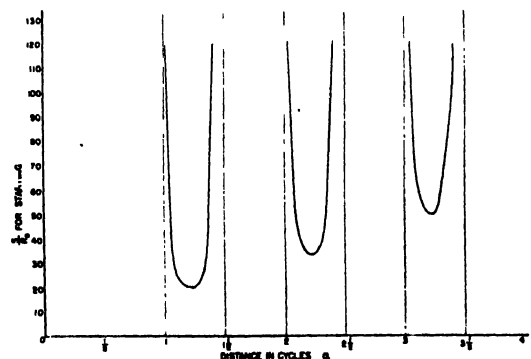


FIG. 12. Starting current function.

By use of a collector electrode on which is placed a voltage which will just stop the slowest electrons leaving the resonator, the efficiency may be increased. According to Müller and Rostas an increase of 7 percent is possible in the first mode.

APPENDIX I. LIMITATION ON EXTENT OF SOLUTION

To find the first electron to be thrown back into the accelerating field we put $\alpha = \partial\alpha/\partial\theta = 0$ from (3b)

$$\frac{\partial\alpha}{\partial\theta} = 1 + \eta \cos \delta - \eta \cos \theta, \quad (16)$$

or

$$\eta = 1/(\cos \theta - \cos \delta). \quad (17)$$

From (17) it is clear that the electron which enters at $\delta = \pi$ is the most likely one to be thrown back. For this electron

$$\eta = 1/(1 + \cos \theta). \quad (18)$$

We may solve for θ from this equation by putting $\alpha = 0$ and $\delta = \pi$ in (3b) which gives

$$(1 - \eta)(\theta - \pi) - \eta \sin \theta = 0. \quad (19)$$

Eliminating η from (18) and (19) gives

$$\tan \theta = \theta - \pi. \quad (20)$$

The first root of this equation, and the only one with physical significance, is $\theta = -1.35$ radians. From (18) we get $\eta_{\max} = 0.823$. For values of η greater than 0.823, some electrons will be thrown back into the accelerating field, and a more complex method of analysis is required. For values of η between 0.5 and 0.823, some electrons will have their direction of motion reversed while traversing the region of the alternating field, but none will be returned into the accelerating field.

APPENDIX II. EQUATION FOR STARTING CURRENT

We indicate here how Eq. (15) for the starting current function R_0/S may be derived. The

problem is exactly the same as stated in Eq. (6), except we are now given the further assumption $\eta \ll 1$. The integral in (6) may now be solved exactly if higher order terms in η are neglected.

The scheme is as follows

$$\bar{\rho} = \frac{1}{2\pi} \int_0^{2\pi} \left[1 - \left(\frac{v}{v_0} \right)^2 \right] d\delta, \quad (21)$$

which from (2b) becomes

$$\bar{\rho} = \frac{\eta}{2\pi} \int_0^{2\pi} [2(\cos \theta - \cos \delta) - \eta(\cos \theta - \cos \delta)^2] d\delta. \quad (22)$$

We must obtain θ as a function of δ from (3b). If $\eta \ll 1$ as a first approximation

$$2\pi\alpha = \theta - \delta \quad \text{or} \quad \theta = 2\pi\alpha + \delta. \quad (23)$$

If this value of θ is substituted in (22) we obtain

$$\bar{\rho} = -2\eta^2 \sin^2 \pi\alpha. \quad (24)$$

This answer, however, has no meaning because we have no information from (3b) regarding η . A more accurate value of θ is obtained from (3b) by substituting for θ from (23) in all terms except the first θ .

$$\theta = 2\pi\alpha + \delta + \eta[\sin(2\pi\alpha + \delta) - \sin \delta - 2\pi\alpha \cos \delta]. \quad (25)$$

When this value for θ is substituted in (22), the integration performed, and all terms in powers of η greater than the squared term dropped, the following formula is obtained for $\bar{\rho}$:

$$\bar{\rho} = 2\eta^2(\pi\alpha \sin 2\pi\alpha - 2 \sin^2 \pi\alpha). \quad (26)$$

Making use of $\bar{\rho} = (k^2/2)(R_0/S)$ and $\eta = k/4\pi\alpha$,

$$\frac{R_0}{S} = \frac{1}{2} \left(\frac{\sin 2\pi\alpha}{2\pi\alpha} - \frac{\sin^2 \pi\alpha}{\pi^2\alpha^2} \right), \quad (27)$$

which is the same as (15). This agrees with the result obtained by Müller and Rostas by a different method.

On the Improvement of Resolution in Electron Diffraction Cameras

JAMES HILLIER AND R. F. BAKER
RCA Laboratories, Princeton, New Jersey
(Received September 6, 1945)

An electron optical system is described in which it is possible to obtain (1) high resolution electron diffraction patterns, (2) shadow electron microscope images, and (3) electron diffraction patterns of selected variable areas of the specimen. The change in mode of operation of the system is accomplished by electrical adjustments without changing the specimen position in any way. A number of phenomena related to the attainment of high resolution diffraction patterns are described. Pertinent details of the experimental apparatus are discussed and some representative results are shown.

IMMEDIATELY after the initial use of the electron diffraction camera as a means of demonstrating the wave properties of an electron beam its further use as a means of analysis was recognized; probably through analogy with the x-ray system. While the special properties of the electron beam have led to the application of the electron diffraction technique to a number of important academic problems concerning the structure of matter, the field of analysis by diffraction methods has continued to be dominated by the older and further developed x-ray diffraction technique. For some time it has been the opinion of the authors that there is no basic reason for this dominance and that if the instrumentation and technique of electron and x-ray methods had been developed equally the former would have definite advantages.

With the introduction of the electron diffraction adapter for the electron microscope¹ some of the advantages of using electronically regulated high voltage supplies and precision electron optics in electron diffraction were realized. However, while this adapter satisfies the desire of the electron microscopist to utilize further the possibilities of his instrument as an analytical tool and provides a means of obtaining high quality electron diffraction patterns, it does not represent, by any means, a complete exploitation of the high precision techniques now available to the electron optical designer.

The electron diffraction adapter has also emphasized an often overlooked but nevertheless important aspect of the electron diffraction method of analysis; namely, the careful correlation of the morphology of the specimen as observed in the light or electron microscope with the nature of the diffraction patterns obtained.

These are only a few of the considerations which have pointed to the necessity of a re-

examination of the electron diffraction technique and to the development of a new instrument which will be described below.

I. THE ELECTRON OPTICS OF THE ELECTRON DIFFRACTION CAMERA

Early electron diffraction cameras followed the

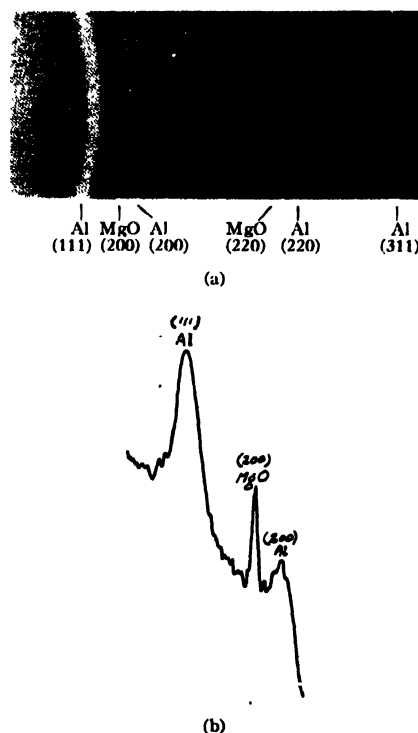


FIG. 1.* (a). An enlargement of a portion of an electron diffraction pattern of a specimen prepared by catching magnesium oxide smoke on a collodion membrane on which a thin aluminum film had been evaporated. This picture demonstrates the difference in natural line widths for different materials. (b). Microphotometer trace of part of the negative used in Fig. 1(a).

* Note: In order to obtain prints of diffraction patterns which were suitable for publication it was found necessary to print the pictures by varying the exposure radially. No interpretation should be made regarding the relative intensities of the lines.

¹ J. Hillier, R. F. Baker, and V. K. Zworykin, *J. App. Phys.* **13**, 571-577 (1942).

same basic design as x-ray cameras, consisting essentially of an electron source, a collimating system arranged to produce a narrow beam of electrons, a specimen, a fluorescent screen, and some means of permanently recording the patterns. It was soon observed that the addition of a focusing lens either ahead of or behind the specimen greatly improved the sharpness of the rings obtained.² The resolving power of this improved type of camera was sufficient to demonstrate natural ring widths in the case of the most commonly used test specimens; namely, evaporated films of aluminum, silver, or gold. It is possible that this was an unfortunate circumstance as it did not provide adequate stimulus for the design of instruments possessing higher resolving power. It is now realized that these particular specimens are somewhat special, most specimens having natural ring widths as much as an order of magnitude smaller and requiring considerably greater resolving power than was obtainable in these earlier instruments. (See Fig. 1.)

It is difficult to assign a value of resolving power to an electron diffraction camera since the obtained resolution is a function of the ring diameter or in other words of the lattice spacing producing a particular spot or ring. If we assume

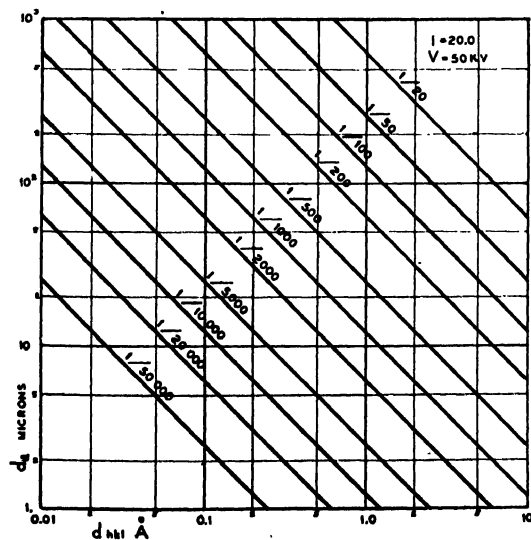


FIG. 2. Shows the maximum center spot size (d_s) which can be tolerated for a given desired resolving power (expressed as a fractional parameter on each curve) and lattice spacing (d_{hkl}).

² G. I. Finch and A. G. Quarrell, *Proc. Roy. Soc. A* **141**, 399 (1933); G. I. Finch, A. G. Quarrell, and H. Wilman, *Trans. Faraday Soc.* **31**, 1051 (1935); G. I. Finch and H. Wilman, *Ergeb. d. exakt. Naturwiss.* **16**, 353 (1937).

TABLE I.

$d_s = 0.01$ cm	$k = 10$ cm ⁻²
$m = 1-2$ (approximately)	$r = 0.025$ cm
$v = 20$ cm	$l = 20$ cm
$f_2 = 10$ cm	

that the focused electron beam at the position of the diffraction pattern consists of a circular disk of illumination of uniform intensity and of radius $d_d/2$, it can be shown that the minimum separation between distinguishable rings is $1.73d_d/2$.^{*} This is assuming further that the rings have no natural width and that the lenses have no aberrations. Using Bragg's law and the geometry of the instrument and assuming that the angle of deflection is small and that there is no lens between the specimen and plate, we obtain an expression for the resolving power of an electron diffraction camera:

$$(\Delta d_{hkl}/d_{hkl}) = (-\Delta D/D) = 1.73d_d d_{hkl}/2l\lambda, \quad (1)$$

where d_{hkl} is the lattice spacing, D the diameter of the corresponding ring, l the distance between the specimen and the photographic plate, and $\lambda = (150/V)^{1/2}$ is the wave-length associated with a beam of V electron volts energy. (See Fig. 2.) In this expression

$$(d_d/2) = (d_s \cdot m/2) + (v/f_2)kr^2 + S, \quad (2)$$

where d_s is the diameter of the electron source, m is the magnification of the lens system, v is the distance between the final lens and the photographic plate, f_2 is the focal length of the final lens, k is the spherical aberration constant of the final lens $f_2 = f_{20}(1 - kr^2)$, and r is the radius of the aperture used in the final lens. The added term S represents the change in spot-size due to the disturbance in the imaging system introduced by the presence of the specimen. This latter effect will be discussed further below. Table I is a typical set of practical values for the early single lens type of camera. In this case the first term of Eq. (2) is the most important except for specimens which charge badly. It amounts to 0.005–0.010 cm and leads to a limiting resolving power in the range 1/70 to 1/170 for $d_{hkl} = 1.5\text{\AA}$

^{*} If, as is usually the case in electron diffraction patterns, the diameter of the ring is much greater than the spot-size, the radial intensity distribution in a ring obtained by uniformly distributing such spots around the circumference of the ring is given by $I_x = C\{(d_d/2)^2 - x^2\}^{1/2}$ where C is an instrumental constant and x is the distance from the center of the distribution measured in the radial direction. The half-value width (i.e., $1.73d_d/2$) of this distribution may be used as a measure of the resolving power providing the rings to be resolved are of equal intensity.

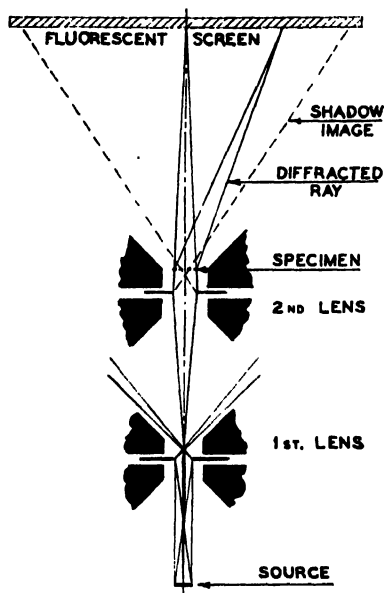


FIG. 3. Schematic diagram of the lens system in the new electron diffraction camera.

and $\lambda = 0.055\text{\AA}$. With the system to be described below, patterns have been obtained which require resolving powers better than $1/400$ to separate rings and as good as $1/5000$ to show the true intensity distribution in those rings, showing that the above resolving powers are inadequate for general work.

Figure 3 shows schematically the lens system of an electron diffraction camera in which a second lens has been added to reduce the size of the final spot. In this case

$$m = m_1 m_2 = 0.01 \times 2.5 = 0.025.$$

Here both the first and second terms of Eq. (2) are important making

$$(d_d/2) = 1.25 \times 10^{-4} + 3.1 \times 10^{-4} + S = 6.7 \times 10^{-4} + S \text{ cm.}$$

Under these conditions the instrument has a limiting resolving power of approximately $1/1000$. By reducing the value of r to 0.0025 cm and the value of d_c to 0.0025 cm the resolving power can be made better than $1/10,000$, a value which is seldom attained in practice due to the appreciable effects of the specimen on the focusing of the spot.

The arrangement of lenses shown in Fig. 3 is the same as that used in the shadow type microscope. This fact leads to the possibility of combining two types of instruments into a very useful analytical tool. The only change in this system necessary to convert it from a diffraction

camera into an electron microscope is a reduction of the focal length of the second lens to place a highly reduced image of the electron source close to the plane of the specimen. In a practical instrument using magnetic lenses this means that it is only necessary to increase the current excitation of the second lens. In one arrangement which has been tried preset shunt resistors can be connected into the coil circuit of the second lens so that the mere snapping of a switch can change the instrument from one type of operation to the other. While the electron microscope images obtained in this manner are not of the high quality of those obtained with a conventional electron microscope, they are quite satisfactory for most purposes. (See Fig. 4.)

The resolving power of the instrument operated as a shadow type electron microscope is given by the expression

$$d_s = (d_e m_1 m_2' / 2) + (\lambda / m_i)^{1/2}, \quad (3)$$

where m_2' is the adjusted value of the magnification of the second lens and m_i is the magnification (electronic) of the image of the specimen. In a practical case the focal length of the second lens is reduced from 10 cm to 0.2 cm (making $m_2' = 0.023$) which leads to a resolving power of the order of $200\text{\AA}$ providing that the electronic magnification of the system is maintained above $10,000\times$. The magnification may be controlled either by further slight adjustment of the power of the second lens or by mechanical translation of the specimen along the optic axis. In this type of electron microscope focusing is neither possible nor necessary; aberrations of the lens system do

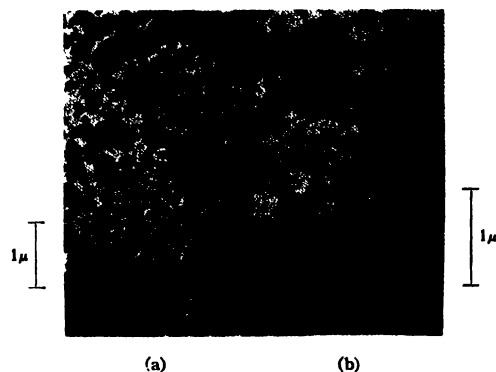


FIG. 4. Comparison between electron micrographs of a clay obtained with the shadow type electron microscope (a) and a conventional magnetic microscope (b). The pictures are not intended to represent the ultimate performance in either case. The shadow micrograph is an enlargement from that obtained in the sequence shown in Fig. 14.

not affect the sharpness of the image but enter as a distortion.

The two modes of operation of the lens system shown in Fig. 3, that is, as a high resolution electron diffraction camera and as a shadow type microscope operating at high magnification, represent the two possible extremes of adjustment. Intermediate to these extremes there is a very interesting continuous range of possible adjustments where the image field consists of a combination of an electron image and an electron diffraction pattern of the same specimen area. The individual diffraction spots are dark field images of the diffracting particles while the central spot is a bright field shadow image of the same particles. Both the diameter of the area of irradiation of the specimen (d_s) and the diameter of the defocused diffraction spot (d_d') vary continuously between these two extremes of adjustment according to the expressions:

$$d_s = 2r \left\{ 1 - \frac{l'}{v_2} \right\} + d_s m_1 m_2' \quad (4)$$

and

$$d_d' = 2r \left\{ \frac{(l+l')}{v_2} - 1 \right\} + d_d \quad (5)$$

where $(1/v_2) = (1/f_2) + (1/u_2)$.

The symbols used are explained in Fig. 5 which is a series of ray diagrams showing the operation of the two lens system as an electron microscope, as an electron diffraction camera, and at an intermediate adjustment. The effect of the spherical aberration of the second lens is neglected in these formula. Using the parameters of the practical case these expressions have been

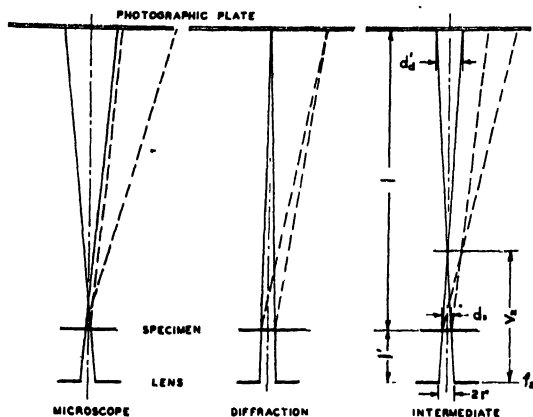


FIG. 5. A series of ray diagrams illustrating the operation of the new camera as a shadow microscope (a), electron diffraction camera (b), and an intermediate adjustment (c).

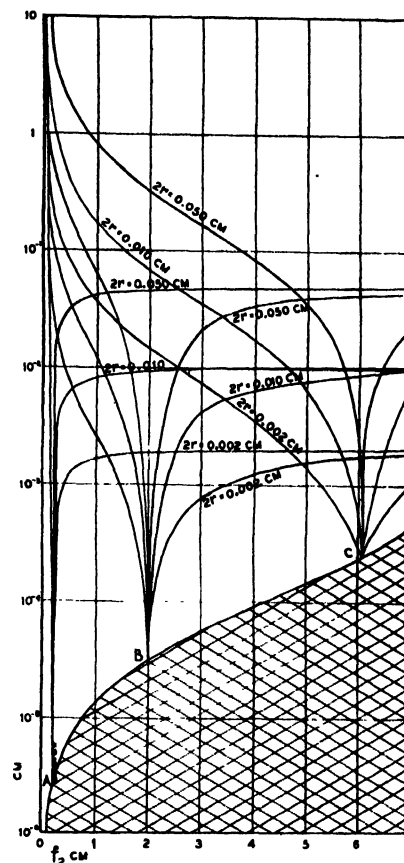


FIG. 6. A set of curves showing a relationship between the diameter of the beam at specified positions in the camera against the focal length of the final lens. The curve, A, B, C , gives the minimum beam cross section. The curves originating at A give the diameter of the area of bombardment of the specimen when situated 2 mm above the second lens. The curves originating at B give the same for the specimen position 2.6 cm from the second lens. Curves originating from C give the diameter of the central spot of the diffraction pattern at the plate position 20.0 cm from the second lens.

calculated for different size second lens apertures and plotted in Fig. 6. From this set of curves it is possible to read off the area of the specimen covered for different adjustments of second lens strength and the spot size (and therefore resolving power) obtained in the corresponding diffraction pattern. The base curve ABC gives the minimum spot size at any plane between the second lens and the photographic plate.

It is obvious that it is not possible to produce a focused electron diffraction pattern of a submicroscopic area of the specimen by electron optical means. It can be accomplished only insofar as the size of the aperture of the final lens can be reduced and insofar as the particular diffraction pattern

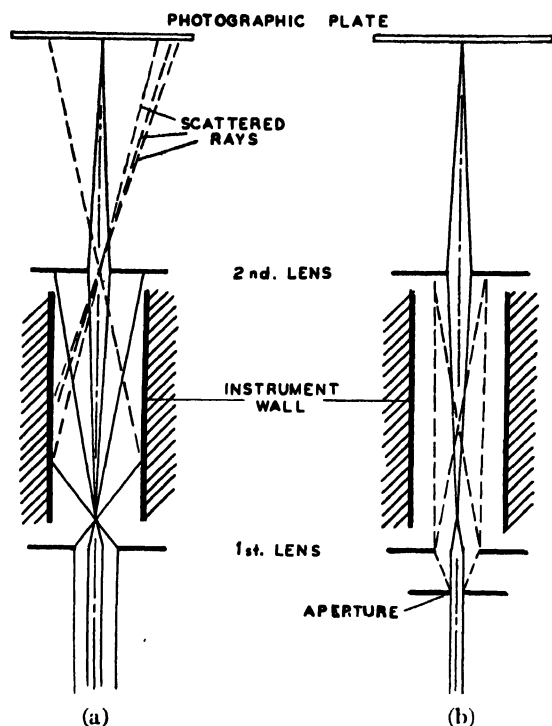


FIG. 7. Ray diagrams showing the origin of spurious patterns due to reflections from apertures and the walls of the system and their elimination by the proper placement of apertures.

encountered can tolerate the required defocusing. For instance, with a 0.002-cm aperture and the focal length of the second lens adjusted to make the image of the source coincide with the specimen plane (point *A* on Fig. 6) the diameter of the area of the specimen irradiated is approximately 230A but the spot diameter in the associated diffraction pattern is 0.2 cm leading to a resolving power of only 1/10 for a lattice spacing of 1.5A. This is hardly sufficient for critical work.

It is possible to reach an advantageous compromise by moving the specimen position further away from the second lens. Such a change is indicated by the series of curves leaving the point *B* on Fig. 6. In this case the specimen has been placed at a distance 2.6 cm from the second lens. The minimum diameter of the area of irradiation is then about 3000A but with the same aperture (i.e., 0.002 cm diam.) the spot diameter in the associated diffraction pattern is 0.015 cm leading to a resolving power of 1/100 for a spacing of 1.5A. This provides a fairly useful arrangement for point diffraction.

As will be seen from the results given below, the operation of this system over its entire range of adjustment is proving very interesting and

fruitful and provides a good illustration of the flexibility of an electron optical system as an analytical tool.

II. ELECTRON DIFFRACTION TECHNIQUES

The establishment of a new instrument as a useful tool and the value of the results obtained from its application depend on the skill and ingenuity displayed both in the operation of the instrument and in the preparation of specimens. This is particularly true in the critical use of the x-ray and electron diffraction techniques. One aid in the attainment of skill in this regard is through an understanding of the possible pitfalls in the method. For this reason some of the difficulties encountered in the use of high resolution electron diffraction are discussed below. The inclusion of these difficulties is not intended as a criticism of the electronic method, it being possible to list equally important difficulties in the corresponding x-ray method.

(a) Vacuum

In electron diffraction work it is necessary to have an extremely good vacuum if dependable results are to be obtained. This is particularly so in the case of surface investigations where contaminants arising from the interaction between the electron beam and the residual vapors in the system are laid down at a rate which quickly masks the original surface. The contaminants are usually organic in nature originating from the materials used in connection with demountable joints, from improperly cleaned metal surfaces or from the oil of the diffusion pumps commonly used.

The full seriousness of this problem has only recently been realized and little has yet been done towards its solution. It is obvious that the problem of designing an instrument which is, at the same time, generally useful and free from all the possible sources of contamination, is an exceedingly difficult one.

(b) Lens Apertures

In an electron diffraction camera of the two-lens type considered above, the pencils of electrons which take part in the formation of the focused spot utilize the full open area of the second lens and hence suffer the corresponding amount of lens aberrations. It is not possible, in this case, to provide an *effective* limitation of the lens aperture, as is done in the electron microscope. Instead it is necessary to provide the lens

with an actual aperture of such diameter that a satisfactory compromise is reached between the intensity of the final spot and the increase in its diameter caused by the aberrations. The proper insertion of an aperture bearing absolutely no surface contamination requires considerable patience and skill.

The arrangement of the lenses leads to another technical difficulty. It involves a long focus lens fitted with a relatively small aperture which is, in effect, an ideal pinhole camera for electrons and will reproduce in the final image plane a faithful image of any part of the instrument between the final lens and the source and in the field of view of the final lens, from which electrons are scattered. This results in a number of spurious markings on the plane of the diffraction pattern. The spurious markings are particularly intense if the electron beam strikes an offending part of the instrument at grazing incidence. Figure 7(a) shows in ray-diagram form the origin of these spurious markings. (See Fig. 8a.) Such markings can be completely eliminated by the proper placement of apertures. In the case of the two lens system being considered it is necessary to place an aperture at the source-side focal point of the first lens. Its diameter must be sufficiently small to prevent electrons striking any part of the system between the first and second lenses except the aperture of the second lens. This is illustrated in Fig. 7(b). By this means no scattering can take place except at the edges of the apertures themselves. The scattering at the edge of the first aperture is focused on the second aperture and is stopped.

Scattering at the aperture of the final lens produces a different effect.* Here the aperture behaves as a secondary source of electrons, producing a shadow image of the entire specimen at low magnification. Figure 8(b) is an example of this type of defect. The use of a clean, conducting, and sharp-edged aperture will minimize this effect.

In considering spurious markings another phenomenon should be mentioned although it has no connection with the aperturing of the system. Occasionally a point of the specimen holder or even a point of the specimen itself which is not in the plane of the diffracting particles will reflect electrons very efficiently. Since such a reflection is the equivalent of a point source close to the specimen the conditions are similar to those of

* The scattered electrons are affected only slightly by the field of the lens.

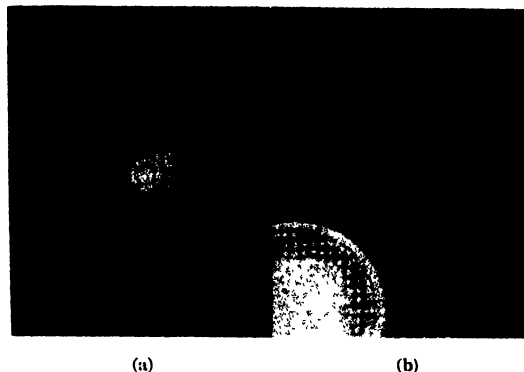


FIG. 8. (a) An electron diffraction pattern of zinc oxide smoke on which can be seen spurious markings due to the reflection from the walls between the two lenses. (b) An electron diffraction pattern of zinc oxide on which is superimposed a low magnification of a shadow image of the supporting wire screen which arises as a result of scattering at the aperture of the final lens.

the shadow type microscope. In this case if the reflection occurs on the lens side of the specimen there will be superimposed on the diffraction a shadow image of part of the specimen. The field of view, intensity, and magnification of this image depend on the accidental placement and size of the reflecting particle relative to the specimen.

(c) Specimen

It is well known that most electron diffraction (and electron microscope) specimens acquire a charge as a result of the bombardment which they undergo. The magnitude of this charge and its distribution on the specimen vary widely according to the physical characteristics of the specimen and the technique of mounting. Except in extreme cases, the steady charged state of the specimen does not disturb the imaging process in the electron microscope. Here the deflection of the individual rays by the charge on the specimen is almost indistinguishable, as far as the microscope objective is concerned, from the scattering which takes place at the specimen and which is a necessary part of the imaging process. In the electron diffraction camera, however, it is not the specimen that is being imaged on the final screen but a distant electron source. Any focusing, deflection, or scattering of the electron beam which takes place at the specimen is not corrected by the lens system and appears as a defocusing, distortion, or diffusion, respectively, of the final spot. Observation of the spot on the fluorescent screen with a 60X magnifier will serve to convince the investigator of the importance of this

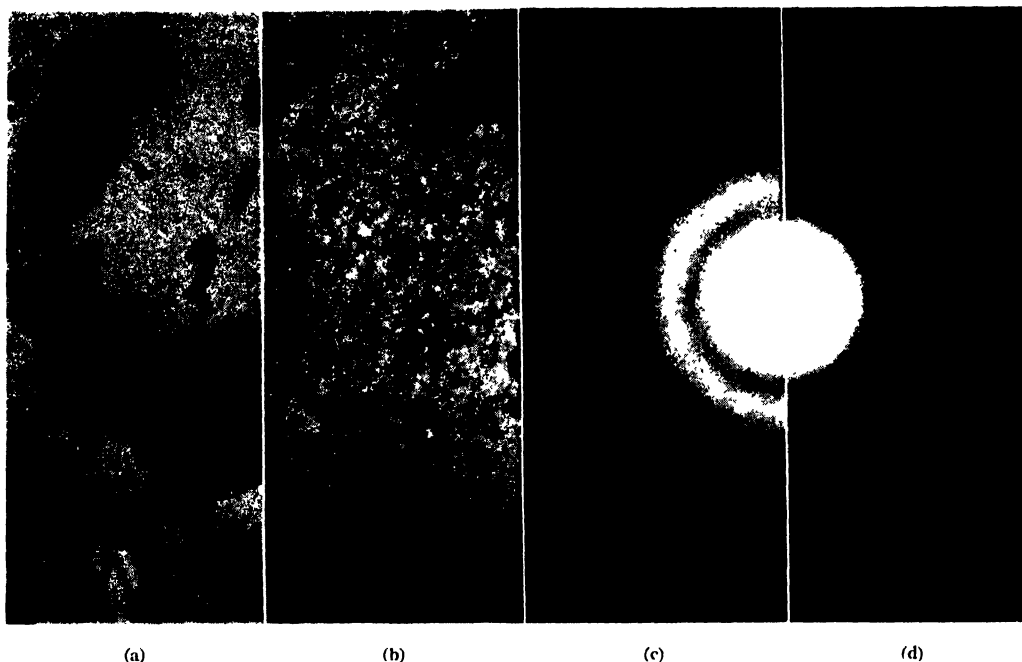


FIG. 9. (a) An electron micrograph of an early dispersion of activated alumina. (b) An electron micrograph of a good dispersion of alumina. (c) An electron diffraction of the poor dispersion. (d) An electron diffraction pattern of the good dispersion of alumina.

effect and provide a means of determining its presence and magnitude.

The effect of charge on specimens for electron diffraction has long been recognized as a difficulty in the case of reflection work and methods of overcoming it have been developed. The authors wish to stress the fact that the effect is present in transmission work and is equally important if high resolution patterns are desired. The effect can be eliminated by the use of an auxiliary low voltage electron "spray." It has also been found that the inclusion of a conductor (connected to ground) in the irradiated area of the specimen will eliminate specimen charge. Proper specimen preparation is also instrumental in reducing the effects of charge.

A complete discussion of the techniques of preparation of the specimen for electron diffraction work is beyond the scope of this paper. However, the importance of the use of the electron microscope for the control of the specimen technique in the case of transmission and reflection diffraction work must be pointed out.

Many powder specimens of industrial importance have a particle size around 1μ . In preparing such specimens for electron diffraction it is essential that the particle size be reduced to a value where the majority of the particles become

at least partly transparent for electrons of the velocity used. The particles must then be mounted in such a way that they do not charge and that in the cross section of the beam there are transparent and properly oriented particles sufficient for the production of a smooth and complete pattern. The mounting must be such that there is either no preferred orientation of the specimen particles or an accurately known orientation. It is obvious that the electron microscope provides an ideal method of observing and controlling these factors. Figure 9 is an example of the effect of proper dispersion on an electron diffraction pattern.

It is unfortunate that the brevity of the preceding paragraph gives such a poor indication of the extreme importance of this aspect of the electron diffraction technique.

III. RESULTS

Using the principles and precautions described in the preceding pages a combination high resolving power electron diffraction camera and shadow type electron microscope has been constructed. Figure 10 is a photograph of the completed instrument. Since the construction of this strictly experimental unit follows the pattern of the experimental electron microscopes designed

in this laboratory only a few significant new features will be described in detail.

As can be seen from the photograph the instrument is mounted on a table-like cabinet in front of the control panel. The beam travels vertically upwards from the gun situated just below the table level and passes successively through the valve chamber, double magnetic lens system, and specimen chamber into the fluorescent screen and photographic chamber. The column is pumped by an oil diffusion pump backed by a mechanical pump. The former pumps through the liquid air trap and large manifold seen to the left of the column and just above the table. Both pumps are mounted below the table in the front part of the cabinet.

One of the most difficult of the problems encountered in the development of the camera was the designing of an electron gun which would give intense single-source illumination and yet at the same time would not be hopelessly critical of adjustment. Owing to the presence of the two small apertures in the lens system the criteria used in the adjustment of an electron microscope gun were not available. A gun satisfying the design requirements has now been developed. It consists of a conventional hair-pin tungsten

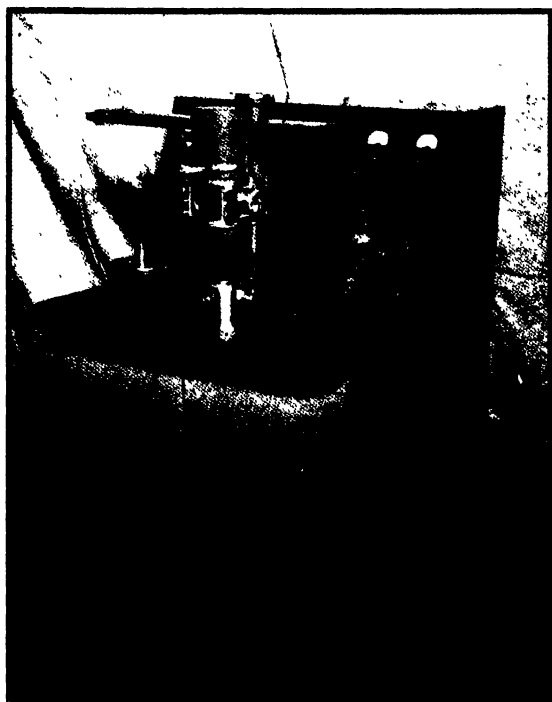


FIG. 10. Photograph of the experimental high resolution electron diffraction camera.

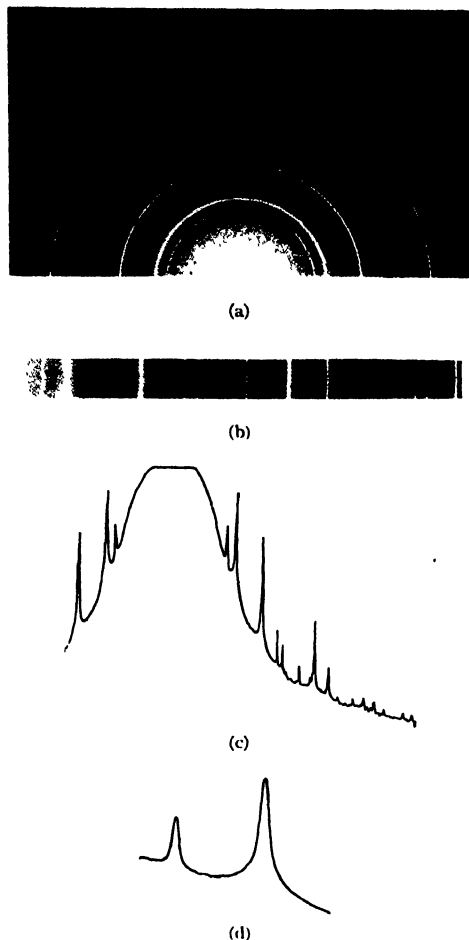


FIG. 11. (a) An electron diffraction pattern of sodium chloride sublimed in a vacuum onto a collodion membrane from a heated filament. (b) An enlargement of part of (a). (c) A microphotometer trace of the negative of the pattern shown in (a). (d) A microphotometer trace of the (111) (200) rings of sodium chloride. The abscissa scale has been magnified in order to show the intensity contours of the rings.

filament axially situated behind, and insulated from, the focusing aperture. The beam current is fed to the filament from the high voltage supply through a high resistance. This arrangement produces a voltage drop which is applied to the focusing aperture. The voltage on the aperture is usually about 500 volts negative relative to the filament and very near the cut-off voltage for the system.

In operation, the filament temperature is set slightly above the value at which the beam current reaches saturation. The combined action of space charge and the dependence of the focusing aperture potential on the beam current serves to stabilize the beam current against small

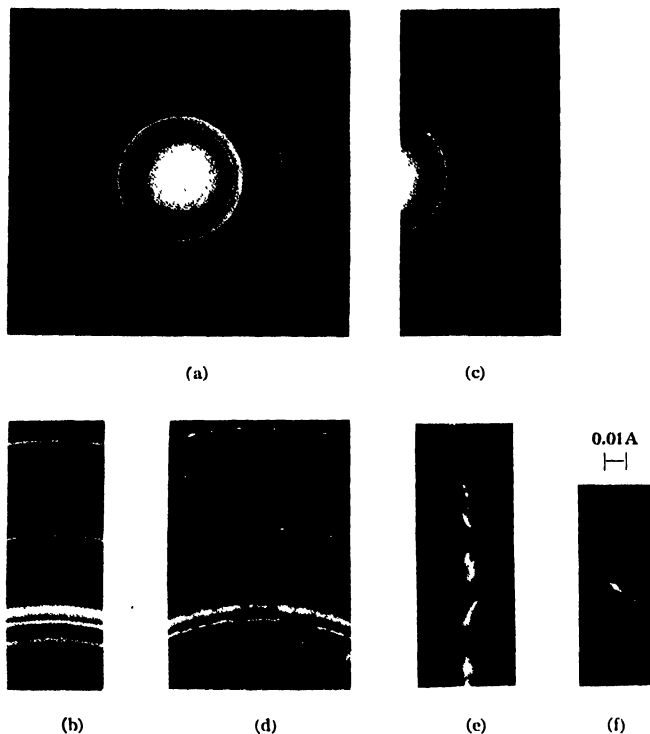


FIG. 12. (a) A normal powder pattern of zinc oxide smoke. (b) An enlargement of part of the pattern in (a). (c) Powder pattern slightly defocused. (d) An enlargement of part of the pattern in (c). The pattern has now become spotty due to the reduction in the irradiated area of the specimen and some of the spots are seen to be elongated due to the fact that they are now relatively low magnification images of the individual zinc oxide spikes. There is an interesting correlation of the individual spots and the line width in the focused pattern in (b). (e) and (f) Still further enlargements of the parts of the pattern in (c). The configuration seen in (f) may be due to a bending of an individual zinc oxide spike. It provides a good demonstration of the resolving power of the system as indicated by the attached scale.

filament current changes and *to center automatically the effective source relative to the focusing aperture*. The source produced is intense and small enough for the most critical electron micro-

scope and diffraction work. The automatic centering of the source relative to the focusing aperture, even for off-center filaments, completely eliminates the necessity for translation and tilt

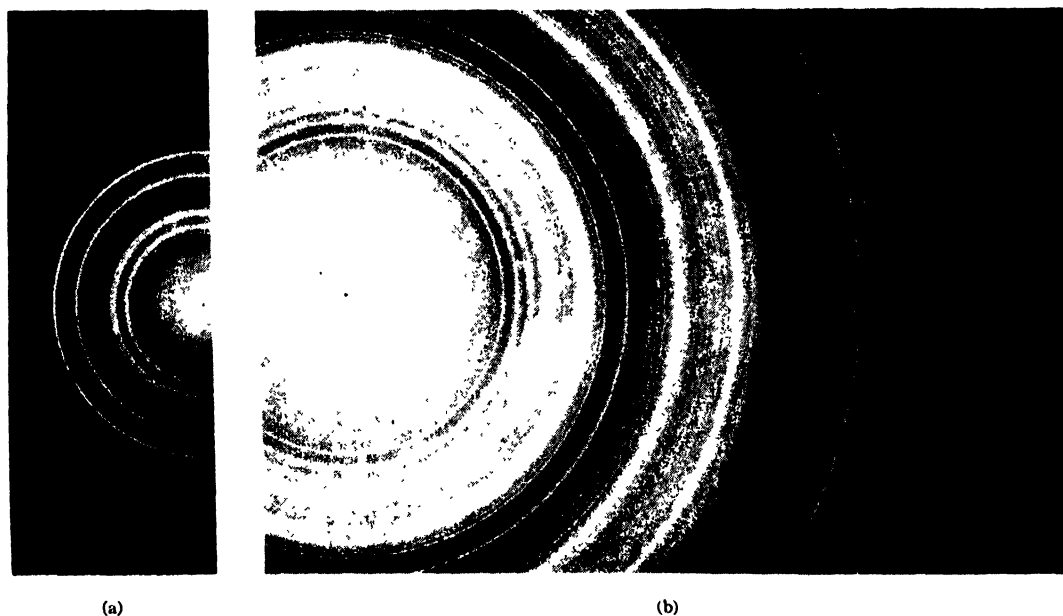
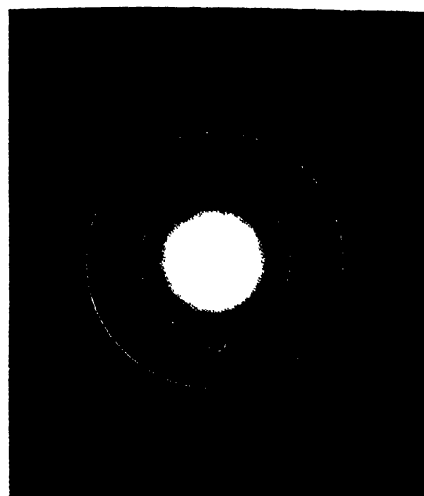
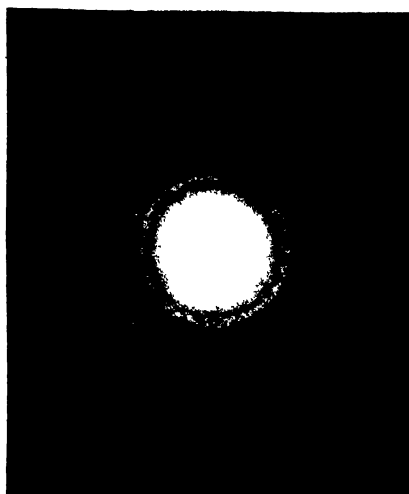


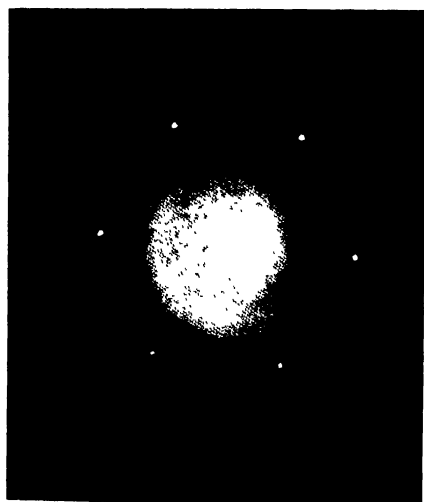
FIG. 13. (a) An electron diffraction pattern demonstrating the resolving power of the new system. The specimen was originally an evaporated film of silver floated from the surface of a rock salt crystal. It was allowed to stand in the atmosphere of the laboratory before the pattern was obtained. Rocksalt, silver, and silver sulphide rings have been identified. (b) An enlargement of (a).



(a)



(b)



(c)



(d)

FIG. 14. A series of pictures of clay showing the complete range of operation of the instrument. (a) A powder pattern. (b) A slightly defocused powder pattern. (c) In this case the instrument is approaching its operation as a shadow microscope. The area of irradiation of the specimen has been reduced until it covers one small group of crystals. At the same time the angular aperture of the irradiation of the specimen has increased so that a number of reflections from different planes are possible. The result is a complete hexagonal pattern. (d) Shadow micrograph of the same specimen.

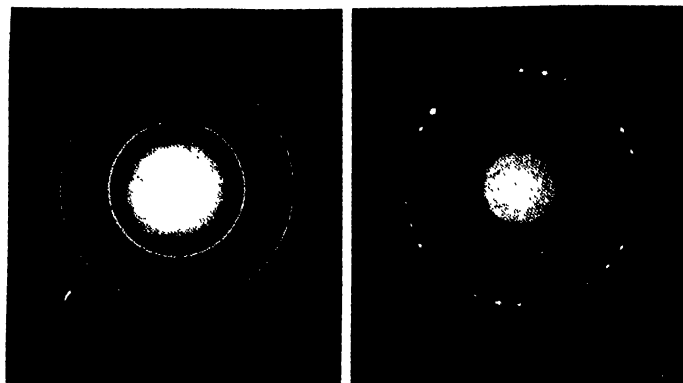
adjustments on the gun. The saturation value of the beam current is determined by the axial separation of the filament tip from the plane of the focusing aperture.

The valve chamber situated between the gun and the lens coil has a valve of new design which when open introduces no obstruction or baffling of the main pumping path. This is accomplished by controlling the valve by a long thin rod sliding through a rubber packing type vacuum seal. When the valve is in the open position the packing can be tightened to prevent any possibility of leakage. Secondary valves are provided for letting air into the system and for by-passing the diffusion pump during prepumping. A large manifold running up the back of the instrument

provides for high speed pumping of the upper sections of the instrument. This manifold is also provided with a liquid air trap.

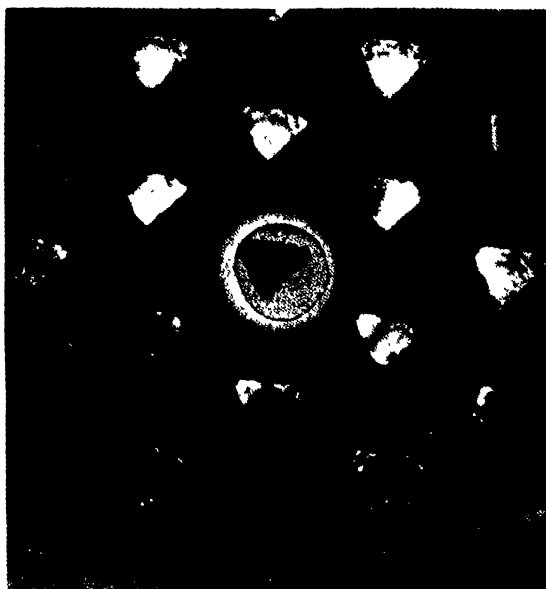
The lens coil and pole pieces are similar to those used in electron microscopes except that the final lens pole pieces are made to produce an asymmetrical axial field distribution. With this system it is not necessary to introduce the specimen into the lens opening. The maximum dimensions of the specimen are then limited only by the size of the specimen chamber.

In the specimen chamber a serious attempt has been made to avoid contamination of the specimen with organic vapors from gasket materials and grease lubrication of the packing seals. It is essentially a double-walled chamber arranged so



(a)

(b)



(c)

FIG. 15. A similar series of pictures of a graphite specimen. (a) Focused powder pattern. (b) Slightly defocused pattern. (c) A considerably defocused pattern in which a single flake was in the irradiation area. Superimposed on the center of this picture is the bright field shadow image of the flake producing the pattern. This image was observed in the instrument simultaneously with the diffraction pattern. It was necessary, however, to make two different exposures owing to the great difference in the intensity of the two types of images. (7000 $\times$)

that there is no direct open path between any gasket or packing seal and the inner chamber. Vapors leaving any of these sources of trouble must diffuse past the manifold liquid air trap before entering the actual specimen chamber.

The viewing system consists of a high resolution fluorescent screen fitted with a forty-five degree front surface mirror and a low-power

compound microscope (30 $\times$). The fluorescent-screen mirror system is mounted on an arm permitting it to be pivoted out of the beam path when photographic records are made.

The photographic chamber uses the standard 2 $\times$ 10 in. plate size and is quite conventional except in that it is fitted with an engraved cassette holder, allowing the accurate positioning of the plate, and with a quick-acting adjustable opening shutter. These features have been introduced to permit the use of the instrument in sequential studies on the treatment of material

which cannot be exposed to air during the experiment. They permit the obtaining of a large number of narrow strip diffraction patterns or micrographs. The power supplies which are quite conventional and which are mounted in the rear of the cabinet consist of a regulated coil current supply with appropriate adjustable shunts for setting the relative powers of the two lenses, a regulated high voltage supply which is adjustable from twenty to fifty kilovolts in five-kilovolt steps, a radiofrequency filament current supply, and the necessary vacuum gauge circuits. All circuits are operated from a line voltage regulator.

The electron optical performance of the system appears to be in good agreement with the predictions of the above mentioned formulae though the total effect of the specimen mounting and charging are by no means thoroughly understood. Some patterns obtained show resolving power exceeding anything in the literature by an order of magnitude. Whether this is a justifiable measure of the performance of the instrument will be better judged from the final appearance of the reproductions in this article. Figures 11 to 15 will serve to illustrate the resolving power of the instrument as a true diffraction camera and its versatility as a probe diffraction camera and shadow microscope.

The authors are greatly indebted to J. M. Morgan who did a large part of the mechanical and electrical design work involved in the development of this instrument.

Applications of Metallic Shadow-Casting to Microscopy^{1,2}

ROBLEY C. WILLIAMS AND RALPH W. G. WYCKOFF
University of Michigan, Ann Arbor, Michigan

(Received July 3, 1945)

The factors which determine image contrast in optical and electron micrographs are discussed in relation to a new metal shadow-casting technique whereby the contrast of images is greatly increased by depositing obliquely a thin film of metal on the microscope preparations. Further advantages of the use of shadow-casting are described, and an estimate is made of the lower limit of size of objects which should be observable by shadow-casting. Examples are given and illustrations are shown of the applications of this technique to the electron micrography of particles of macromolecular dimensions, of replicas of such particles, and of surface replicas prepared in several ways.

INTRODUCTION

A GOOD micrograph gives clear and accurate information about the size, shape, and structure of the object from which it is derived. Whether one is dealing with microscopy involving light or electrons, the limit to what can be seen in a micrograph is in the end determined by the resolving power of the lens system and by the contrast between adjacent elementary regions of the object. Most of the contrast in the optical microscope results from differences in absorption of light within the preparation. Many important classes of objects, particularly those of biological origin, have in themselves little variation in absorption for visible light, but this restriction on microscopic examination, of course, has been in large measure eliminated through the development of elaborate methods of differential staining using dye-stuffs and other coloring agents, and to a lesser degree through the use of ultraviolet light. Under these circumstances the magnification, and the smallness of detail that can be observed with the optical microscope, are set by the attainable resolving power, i.e., by the wave-length of the light used.

Conditions are somewhat different in electron micrography at the present stage of its development. Contrast in an object illuminated by a beam of electrons is not due to differential absorption but rather to an unequal amount of scattering of electrons at different points in the preparation. The visibility of details of an object

resulting from the contrast they introduce into an electron micrograph is thus determined largely by the amounts of matter contained in these details. Such amounts of matter may be so small that, especially for objects resting on a supporting membrane, insufficient contrast rather than inadequate resolving power of the microscope fixes the lower limit of visibility. The present paper is devoted to a discussion of some of these limitations arising from insufficient contrast, and to a description of a technique by which the contrast in specimens for microscopy can be greatly enhanced.

The image responsible for an electron micrograph is formed by focusing the electrons that penetrate the specimen and its substrate. The contrast in the final image is a result of failure of the lenses to focus accurately those electrons scattered through relatively large angles while passing through specimen and substrate. If these widely scattered electrons are allowed to strike the photographic plate they cause a generally distributed fogging. Maximum of contrast is therefore achieved by employing a small objective aperture which will exclude such widely deflected electrons and pass only those that can be focused with precision by the objective and projector lenses. All the electron micrographs described in this paper have been made with an RCA Type EMB instrument equipped with an objective aperture.

Simple considerations make evident the lower limit of size of biological and other objects of low density which can be recorded by direct photography with the electron microscope. One can, for example, scarcely expect to observe

¹From the Department of Physics, University of Michigan.

²This research was supported in part by the Evaporated Metals Films Corporation, Ithaca, New York.

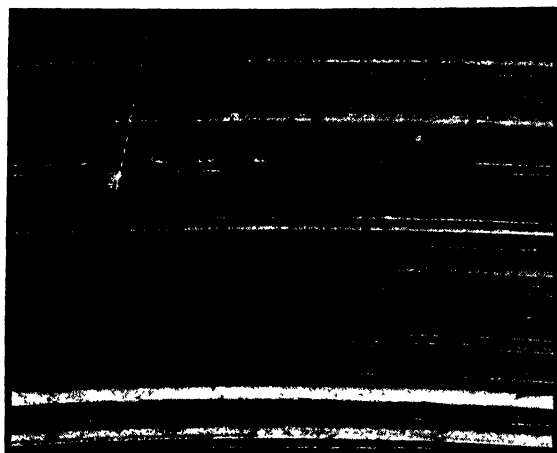


FIG. 1. An aluminized collodion replica of a machined brass surface. 10X

directly the details of surface structure on a bacterium. Such an organism will have a thickness of not less than several tenths of a micron and it will be supported on a film of collodion or Formvar³ which supplies additional scattering matter. It is evident that surface details which provide variations in thickness of amounts less than, say, 0.02 micron will introduce totally imperceptible differences in scattering and contrast in micrographs. In the range of particles smaller than bacteria it is clear that none can be detected if its scattering is insignificant compared with that of the substrate on which it rests. All organic substances have about the same low scattering for electrons, and experience shows that under the most favorable conditions it is hard to detect the existence of particles supported by a substrate and having diameters less than about 0.01 micron. By the same reasoning it is impossible to form an accurate idea of the shapes of much bigger particles. Unless they are flat or are enclosed in thick membranes, the outlines of even relatively large organic objects must be diffuse on direct electron micrography, since at some point within their geometrical edge they will have thinned down to such an extent that the scattering is no longer considerable compared with that of the substrate.

The same kind of diffuseness of fine detail is observed in an electron micrograph of a collodion

or Formvar⁴ replica of an etched metal or other rough surface, and it has usually been concluded⁵ from this that such replicas are not sharp. The work described in this paper shows that such simple replicas are faithful to the surface from which they are stripped and that their apparent lack of sharpness arises from the insufficient contrast introduced by their small scale variations in thickness and scattering. This conclusion has great practical importance because replicas made with polymers in solution are especially convenient to prepare, and can be stripped from many surfaces and materials that cannot be subjected to the high temperatures and pressures involved in making polystyrene⁵ and other molded replicas.

THE SHADOW-CASTING TECHNIQUE

In a previous paper⁶ we have described a technique for enhancing contrast in electron microscopic preparations which we have called shadow-casting. Briefly it consists in coating the specimen obliquely with a very thin film of metal prior to the microscopic examination. The coating is accomplished in an auxiliary vacuum chamber by means of a filament containing the metal to be deposited. The effect of the oblique deposition is to cause the higher elevations in the specimen to cast total or partial metallic "shadows" on the sides or slopes away from the filament. When the shadow-cast specimen is photographed, it yields a negative that shows light and dark areas whose degree of darkness, as a measure of the thickness of deposited metal, is a function of the angle between the tangent planes of the areas and the direction to the filament during evaporation. Such a negative, or negative print, of a shadow-cast specimen gives the same impression that is produced by a picture taken with light reflected from an obliquely illuminated landscape.

Any substance that does not exhibit structure of its own in thin films and does not migrate after deposition on the specimen could be used for shadow-casting, but it is clear that only the

⁴ V. J. Schaefer and D. Harker, *J. App. Phys.* **13**, 427 (1942).

⁵ R. D. Heidenreich and V. G. Peck, *J. App. Phys.* **14**, 23 (1943); R. D. Heidenreich, *ibid.* **14**, 312 (1943).

⁶ R. C. Williams and R. W. G. Wyckoff, *J. App. Phys.* **15**, 712 (1944).

³ Formvar 15-95 is a polyvinyl formal supplied by the Shawinigan Products Corporation, New York, New York.

heavier elements can provide the extremely thin films of great scattering power that are needed in electron microscopy. Very thin films are generally desirable since they cause a minimum of alteration in the geometrical outline of the specimen on which they are deposited. Chromium yields a notably structure-free film and is ideal for shadowing relatively large objects such as bacteria,⁷ but the considerable thickness of the requisite film (*ca.* 70Å) makes it unsuitable for delineating objects of macromolecular dimensions. For such smaller objects gold⁸ in a computed thickness of *ca.* 8Å on the plane of the substrate gives satisfactory results.

Shadow-casting has a number of desirable consequences, first of which is that the heights or thicknesses of discrete objects that are not wholly flat are readily estimated from the lengths of the shadows⁹ they have cast. Relative heights can of course be obtained in other ways, for instance by stereoscopy,¹⁰ but these other procedures are more laborious, and with them it is hard to be sure which areas are high and which are low. Shadow-casting brings out in great relief the surface contours of the specimen by enhancing the scattering power, or optical effectiveness, over the entire preparation and by causing this increase to vary with the contours and to concentrate in the surface layers of the specimen. The type of detail brought out by shadowing can be varied by altering the thickness of the evaporated layer; for example if it is so thick that its optical effectiveness exceeds that of the rest of the object, it will portray almost exclusively the details of the surface. Shadowing also affords contrast between areas whose elevations in the specimen are similar but whose slopes are different, or between areas whose slopes are the same but are separated by regions of different slopes. Furthermore it makes clearly evident objects so small or thin that by themselves they would not produce a minimum contrast against the background of the rest of the specimen. Experience indicates that an ob-

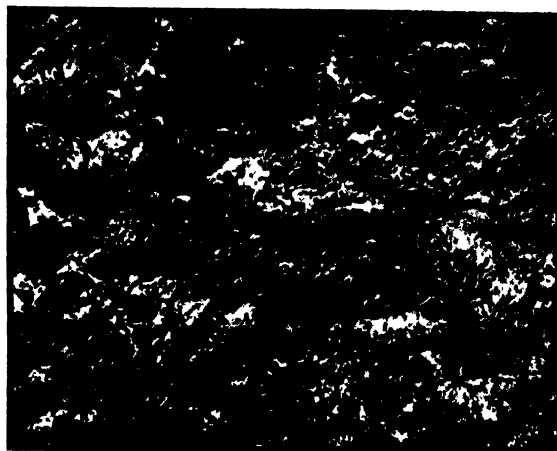


FIG. 2. An aluminized "Faxfilm" replica from an anodically oxidized high speed drill. 500X.

ject even smaller than the lower limit of resolution of the electron microscope as it now exists can be made to cast a partial or total shadow at some highly oblique angle of illumination and that it can be detected by the existence of this shadow.

SHADOW-CASTING FOR OPTICAL MICROSCOPY

Shadow-casting will increase the contrast on replicas which are to be studied and photographed with the optical microscope, as well as with the electron microscope. Lower magnifications will commonly be employed for examination with visible light, and the metallic film that produces the shadows need not be unusually fine grained. As a consequence of this freedom of choice of film, aluminum of such a thickness that it is partially opaque to visible light can be used as a convenient shadowing metal for such "visual" replicas. We have made satisfactory visual replicas starting with dissolved collodion or Formvar, or with the commercial replica film process known as "Faxfilm."¹¹ Figure 1 is a photograph of an aluminum-shadowed collodion replica of a machined brass surface, magnified ten times. Figure 2 shows an aluminized "Faxfilm" replica of an anodically oxidized surface of a high speed drill, magnified 500 times. Shadowing brings out much fine detail in this "Faxfilm" replica that could not be seen otherwise. Figure 3 is an aluminized collodion replica

⁷ R. C. Williams and R. W. G. Wyckoff, *Proc. Soc. Exp. Biol. Med.* **59**, 265 (1945).

⁸ R. C. Williams and R. W. G. Wyckoff, *Science* **101**, 594 (1945).

⁹ H. O. Müller, *Kolloid Zeits.* **99**, 6 (1942).

¹⁰ See R. D. Heidenreich and L. A. Matheson, *J. App. Phys.* **15**, 423 (1944), for references.

¹¹ Faxfilm, R. D. McDill, Cleveland, Ohio.

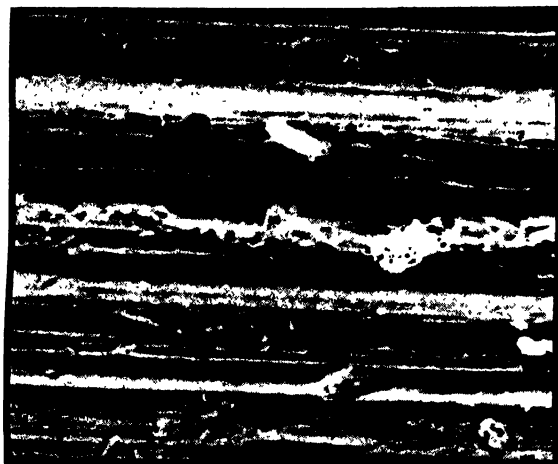


FIG. 3. An aluminized collodion replica of a machined steel surface. 1000X.

of a machined steel surface at a magnification of 1000 times. The contrast in such shadowed objects is so high that the condenser of the microscope can be employed at the full aperture necessary for maximum resolution.

SHADOW-CASTING FOR ELECTRON MICROSCOPY

1. Replicas of Surfaces

The electron microscope has proved to be especially useful in examining replicas of surfaces of many types of objects. Several methods of preparing these replicas have been described.¹² Two of them, however, are in greatest current use: the Formvar process, as developed by Schaefer and Harker, and the polystyrene-silica method of Heidenreich and Peck.

As stated above, a Formvar, or a collodion, stripping lacks sharpness when viewed at high magnification in the electron microscope, and this generally has been ascribed to its intrinsic lack of sharpness of outline and faithfulness to the surface from which it is drawn. We have found on the contrary that replicas of this sort appear geometrically sharp after being shadowed, and that their unsharpness when photographed before metal deposition must be ascribed to a lack of contrast between regions of nearly the same thickness.

¹² H. Mahl, *Zeits. f. tech. Physik* 21, 17 (1940); *ibid.* 22, 33 (1941); *Metalwirtschaft* 19, 1082 (1940); V. K. Zworykin and E. G. Ramberg, *J. App. Phys.* 12, 692 (1941); V. J. Schaefer and D. Harker, see reference 4; R. D. Heidenreich and V. G. Peck, see reference 5.

Polystyrene-silica replicas exhibit good contrast but probably not for the same reason that our obliquely metal-shadowed films do. Heidenreich has suggested¹³ that silica deposits upon a polystyrene base in such a way as to form a layer of uniform thickness measured normal to the elements of the surface. Hence, as seen by the electron beam, regions of the silica film which are sharply tilted to the general plane of the replica will be effectively thicker than regions parallel to the plane. Highly tilted regions will consequently appear dark on positive pictures and flat regions will appear light. If this hypothesis is correct it makes detailed interpretation of the micrographs difficult, since only changes in elevation can be recognized. Stereoscopic photography can profitably be used in understanding these polystyrene-silica replicas, but it does not ordinarily resolve the fundamental ambiguity in them between what is high and what is low. This interpretation of the form of the silica film is supported indirectly by our observation that a replica film made by depositing a uniform film of chromium in a good vacuum at normal incidence upon a polystyrene replica cast, unlike a

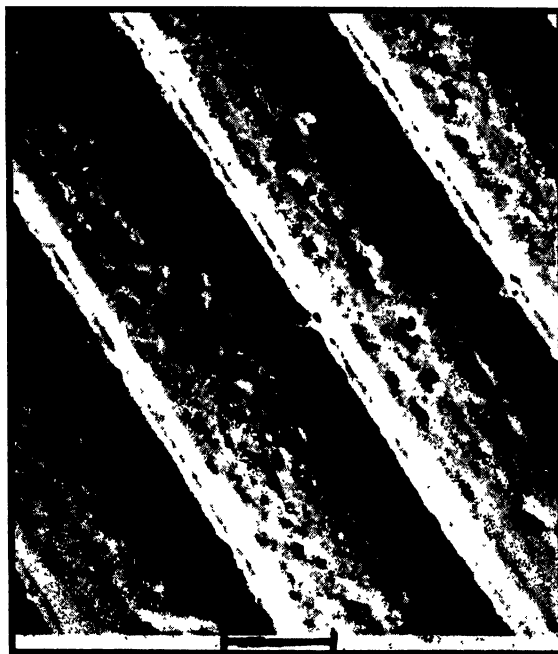


FIG. 4. An all-metal chromium-shadowed replica of a glass diffraction grating.

¹³ R. D. Heidenreich, see reference 5.

deposited silica film, introduces no contrast in the micrographs made of it. In this case the metal, which does not migrate after deposition, can be expected to form a film varying in thickness over the contours of the surface according to the cosine law, thereby producing thinner films on highly tilted areas as measured normal to the area. As a result the effective thickness of the film as seen by the electron beam will be sensibly constant over the entire replica and the result will be an almost complete lack of contrast.

We have found a number of different procedures satisfactory for preparing replicas, depending on the surface to be reproduced and the use to which the replica and its photograph are to be put. At one stage or another, however, we invariably include a high vacuum deposition of a thin metallic film at a known oblique angle.

a. All-metal replicas can be made of glass and of many crystalline surfaces. One way to make them involves evaporation from three filaments. First there is deposited upon the original surface, at normal incidence, an extremely thin (*ca.* 20Å) film of acid-, or water-, soluble material such as copper or NaCl. This is followed by a film of chromium 100Å thick evaporated also at normal incidence to provide body to the replica. The third film, which introduces contrast, can be another 100Å of chromium deposited obliquely.

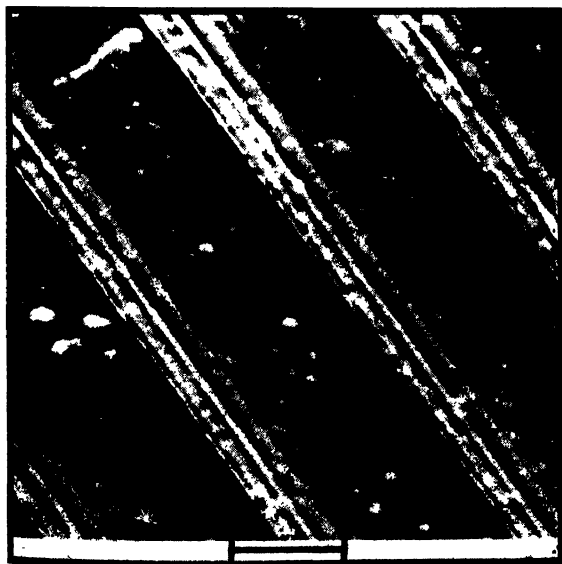


FIG. 5. A gold-shadowed collodion replica of the diffraction grating of Fig. 4.

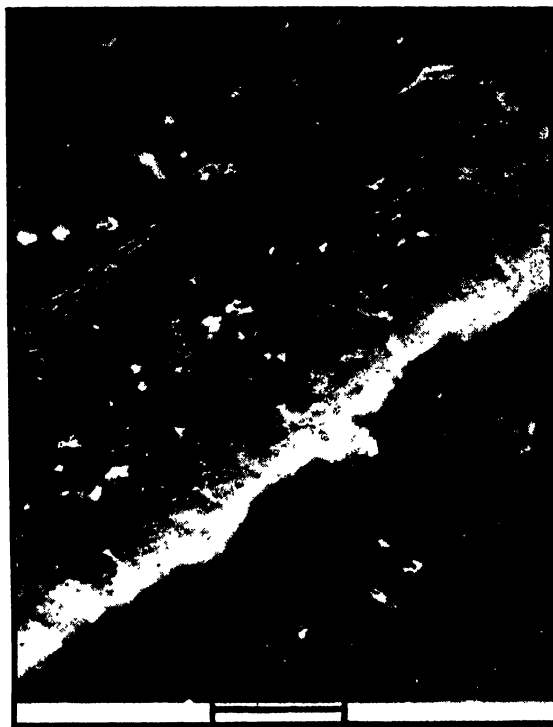


FIG. 6. A chromium-shadowed collodion replica of a surface of a used roller bearing.

The specimen with its attached films is then immersed in dilute acid or in water in order to float off the coherent film of chromium, and this is then picked up on a microscope screen by "fishing." Figure 4 shows such an all-metal replica of a 15,000 line-per-inch glass diffraction grating.¹⁴ The intermediate soluble layer has in our experience introduced minute pits that make this type of replica unsuitable for magnifications exceeding *ca.* 20,000 times. The all-metal replica is chiefly useful when, as for calibration of magnification, one wants a specimen that can be relied on not to expand or shrink perceptibly while being made, and that does not charge up and distort under the impact of the electron beam.

b. Shadow-cast collodion replicas¹⁵ are especially easy to make and to use, and with a fast evaporating unit the time elapsed in passing

¹⁴ All the electron micrographs of this paper are from negative prints of intermediate contact positives on lantern slide emulsion. The scale of magnification in each instance is indicated by the one micron scale underneath the photograph.

¹⁵ L. Thomassen, R. C. Williams, and R. W. G. Wyckoff, *Rev. Sci. Inst.* 16, 155 (1945).



FIG. 7. A chromium-shadowed collodion replica of a scratched glass slide.

from the original specimen to the finished replica can be as short as 15 minutes. The average time required to make such a replica, or a corresponding one with Formvar in place of collodion, is further reduced since a great many strippings can be shadowed at one time. For the sake of comparison between the all-metal and the collodion-base processes, there is reproduced in Fig. 5 a collodion-gold (10A deposited at a four-to-one angle) replica of the same glass diffraction grating employed to make Fig. 4.

Figures 6 to 11 are additional examples of electron micrographs of replicas prepared by this collodion-metal process. All the collodion strippings for these replicas have been shadow-cast at angles of five-to-one ($\arctan 1/5$) either with a film of chromium *ca.* 80A thick or with *ca.* 8A of gold.

Figure 6 is a chromium-shadowed replica of the inner surface of the outer cone of a used roller bearing. This picture illustrates a great advantage of the process: it can be employed with many kinds of intact objects whose shapes, sizes, and locations are such that molded replicas could not be drawn from them.

Figure 7 is a replica of an accidental scratch on a glass microscope slide. It shows the unusual smoothness of the undamaged glass surface as

well as the large amount of small scale tearing and flowing in the scratch.

Figure 8 is a gold-shadowed replica of the surface of a diamond polished parallel to a (110) plane. The approximately circular areas about one micron across correspond to elevations on the polished surface, and in this field a sharp minute pit can be seen near the center of each elevation. The residual polish marks are clearly seen running parallel across the surface of the diamond; the depth of the smallest of these is *ca.* 50A.

Figures 9-11 are gold-shadowed collodion replicas of the polished and etched surfaces of pieces of steel, of 60-40 brass and of magnesium. They illustrate the ease with which determinations can be made of relative heights and depths in such specimens. This is particularly evident in Fig. 11 where four distinct levels can be seen on the surface of the replica. The center is a high triangular peak surrounded by a slightly raised region of roughly triangular shape; around this



FIG. 8. A gold-shadowed collodion replica of a polished diamond face.

is the general level of the replica surface which is partially covered with nearly circular pits.

c. Replicas can also be made by depositing first a metal film obliquely on the surface of the specimen. If the film is sufficiently thick and the specimen has an appropriate surface it can sometimes be stripped directly to yield a simple all-metal replica. Otherwise the film can often be backed up by forming on it a supporting collodion film and stripping the two together from the underlying specimen. This process does not appear to be applicable to most etched metallic surfaces but, as will be indicated at a later

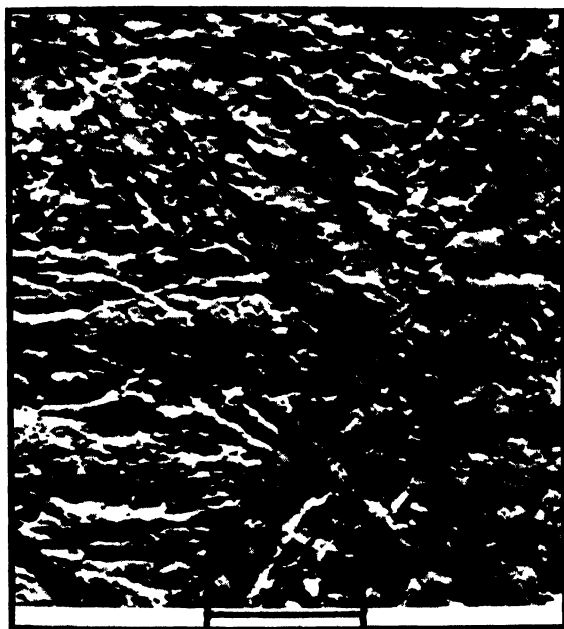


FIG. 9. A gold-shadowed collodion replica of an etched steel surface.

point, it provides a very valuable way of studying macromolecules and other minute particles that can be deposited on a glass surface prior to being shadowed.

2. Specimens mounted on Substrates

The advantages of shadow-casting discussed earlier in this paper are all applicable to electron microscopic preparations in which the object of study is mounted for examination on a collodion or Formvar substrate. Figures 12 to 14, and the photograph on the cover of this Journal, have been chosen to illustrate the applications that



FIG. 10. A gold-shadowed collodion replica of an etched surface of 60-40 brass.



FIG. 11. A gold-shadowed collodion replica of a piece of magnesium etched with salt water.

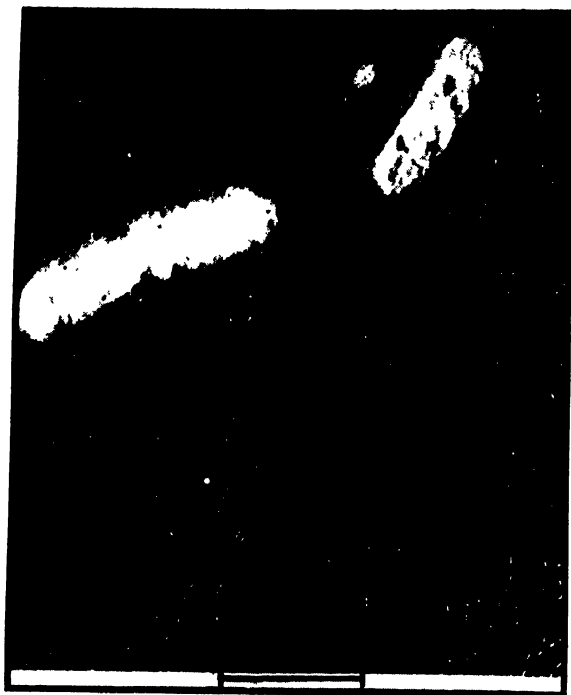


FIG. 12. Chromium-shadowed typhoid bacilli.

have been made of shadow-casting to the observation of biological objects of very different sizes when mounted on collodion.

Bacteria can conveniently be used to demonstrate the kind of new information that can be brought out by shadow-casting relatively large objects. Figure 12 is a photograph of chromium-shadowed typhoid bacilli which have been given an exposure in the electron microscope just sufficient to show properly the collodion background and such small details on it as the bacterial flagella, both intact and fragmented. Though the organisms themselves are underexposed it is nevertheless possible to see something of their surface structure. Figure 13 contains a group of old cells from a culture of *Staphylococcus aureus*. They have been shadowed with chromium and exposed in the electron microscope long enough to bring out much surface detail that is not otherwise seen. The way in which shadow-casting supplements the direct electron micrography of large biological objects such as bacteria is evident from these pictures and others that have been published.⁷ Since it concentrates scattering in the superficial layers of the object, it has necessarily

the general effect of hiding any inner structural details that may exist. Hence, for the most complete photographic survey of such objects three kinds of photographs should be made: micrographs of the unshadowed preparation, brief exposures, and prolonged exposures of the preparation after shadowing. In this way one obtains information regarding inner structure, structure around the periphery and adjacent to the object, and details of its surface.

In photographs we have published elsewhere of influenza virus particles¹⁶ and of tobacco mosaic virus fibrils^{8,16} we have demonstrated the application of the shadow technique to biological entities too small to be seen by optical microscopy. We have also shown that the technique permits the unequivocal photography of the elementary molecular particles of such macromolecular substances as the hemocyanins.¹⁷ Figure 14 is a gold-shadowed preparation (at ten-to-one angle) of a very dilute ultracentrifugally purified solution of the hemocyanin from the horseshoe crab, dried on a collodion sub-

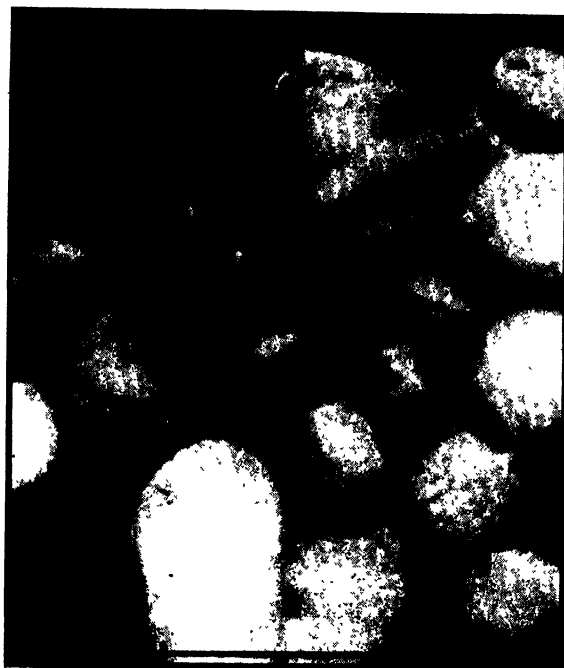


FIG. 13. Chromium-shadowed cells from an old culture of *Staphylococcus aureus*.

¹⁶ R. C. Williams and R. W. G. Wyckoff, Proc. Soc. Exp. Biol. Med. **58**, 265 (1945).

¹⁷ R. C. Williams and R. W. G. Wyckoff, Nature **156**, 68 (1945).

strate. A few molecules distributed singly and in small clusters can be seen lying on the sharply granular collodion background. These molecules, while clearly visible and casting well-defined shadows, are not much bigger than the individual collodion granules; and it would be hard to distinguish much smaller particles from the collodion on which they rest. In other words, the fine structure of the supporting collodion membrane itself sets a lower limit to the size of particles which can be mounted on it and investigated.

The picture on the cover of this Journal is another example of the photography of macromolecules supported on a collodion membrane and shadowed with gold (deposited at a five-to-one angle). The bushy stunt virus,¹⁸ whose elementary particles are the small spherical bodies covering most of the field of the micrograph, is known to crystallize readily from



FIG. 14. Gold-shadowed hemocyanin molecules on a collodion substrate.

¹⁸ W. C. Price, R. C. Williams, and R. W. G. Wyckoff, *Science* 102, 277 (1945); *Arch. Biochem.* In press.



FIG. 15. A gold replica of hemocyanin molecules on a glass surface. For comparison with Fig. 14.

purified solution, and this tendency of the molecules to form regular arrays is very apparent in this figure. Other photographs demonstrate how similarly oriented molecular layers pile one on top of another to produce the three-dimensional symmetrical arrangements of molecules that constitute a true crystal and are responsible for its distinctive properties.

3. Preshadowed Replicas

It is evident from a photograph such as Fig. 14 that, in order to investigate molecules the size of the granules seen in a collodion film, the molecules should be on a smoother substrate when shadowed. We have found^{8,17} that one way to do this is to employ a technique which is essentially that mentioned above under 1c. A drop of the solution to be studied is first placed on a piece of clean, polished glass and allowed to evaporate to dryness. The glass with the desiccated material adhering to it is next shadowed obliquely with *ca.* 8Å of gold. Since such a thin film is too fragile to be handled directly, it is backed by the usual thin film of collodion or Formvar. The polymer and the metal films are then stripped together from the glass and mounted on a specimen screen. Figure 15 shows a micrograph

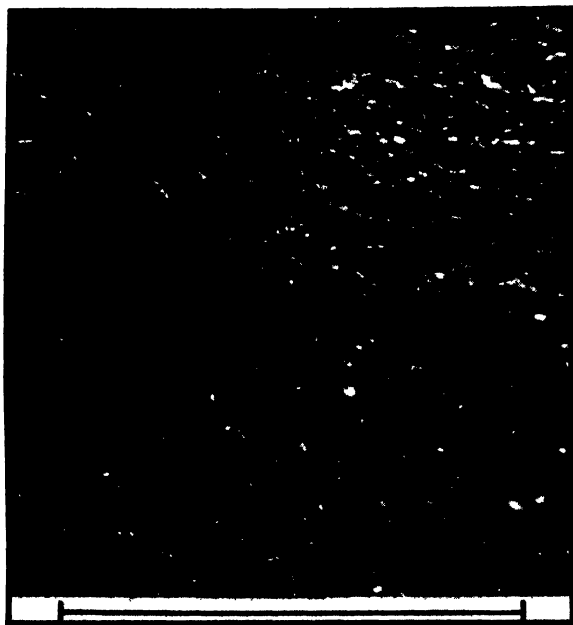


FIG. 16. A gold replica of the cellulose molecular fragments from collodion, on a glass surface.

of a preshadowed replica of limulus hemocyanin molecules for comparison with Fig. 14. The molecules themselves appear equally well defined and of the same size on both photographs, but since the collodion was not metal-shadowed in the replica preparation for Fig. 15, its granular fine structure is not brought out. Obviously, far smaller particles can be distinguished and photographed on the smooth background of Fig. 15 than on a pebbly one such as that of Fig. 14.

Examination of a photograph of this type immediately raises the question of whether or not the particles that are being studied are removed with the stripped films. When dealing with objects as large as bacteria we have found that they sometimes are retained in the replica and sometimes left on the glass; which is the case in any particular instance is obvious from the opacity of the image in the electron microscope. With very small particles like the molecules of hemocyanin it makes no practical difference which is true because the electron scattering power of the metal film greatly exceeds that of the particles themselves.

Preshadowing also makes it possible to study the intimate structure of collodion and of other highly polymerized substances which must be

dissolved in organic solvents that would attack the materials ordinarily used as substrates. This application is illustrated in Figs. 16 and 17. Figure 16 shows the nitrated cellulose fragments that make up collodion, as photographed in this way. The object for the left half of Fig. 16 was collodion very greatly diluted with amyl acetate and spread on glass to dry. The molecular fragments are thread-like aggregates, of various lengths, of the same granules that cover the entire surface of Fig. 14. For the right half of Fig. 16 the collodion solution was not quite so dilute. It shows how, in the formation of a continuous film, these beaded threads pack together with their long axes parallel. At the top of this figure the cellulose threads are widely spaced; towards the bottom the film is nearly continuous.

A preshadowed gold replica (at a ten-to-one angle) of polystyrene dissolved in ethyl bromide has been photographed to yield Fig. 17. The fragments strewn over the glass background vary much more in size than do those seen in the collodion photographs and some of them are of an

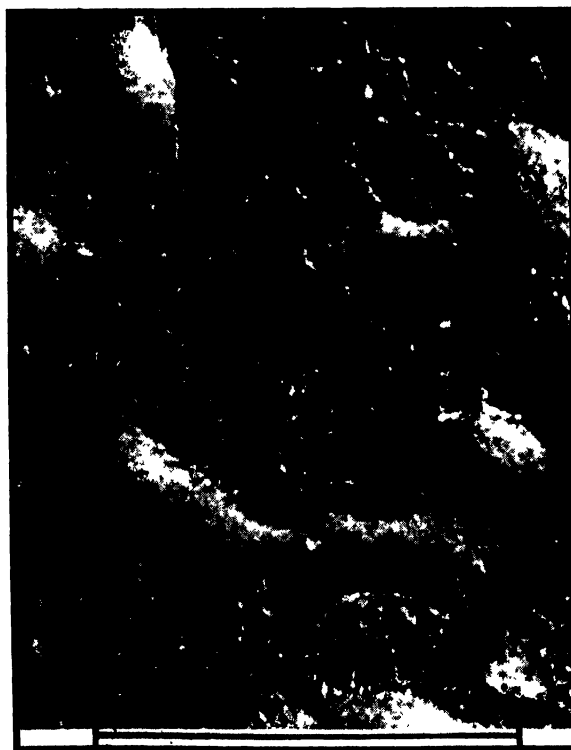


FIG. 17. A gold replica of polystyrene molecular fragments on a glass surface.

exceedingly small cross section. Measurements of shadow lengths indicate that some are not more than 15Å high. Though it may not be too apparent in the print, inspection of the original negatives shows that the large droplets on this and other photographs are aggregates of the smaller fragments such as those of Fig. 17. It is clear from our experience with many pre-shadowed replicas that the minimum of particle size which can be detected on them is now set

by the smoothness and cleanliness of the glass surface employed. At least for the very small areas needed for this type of work, glass can be found and cleaned so as to be free from abrupt variations of surface elevation within an amount not exceeding five to 10Å, as judged by observed shadow lengths, and hence it can be concluded that electron microscopy makes it feasible to detect the existence of sharply defined objects of this range of height.

The Electrical Resistance of Iron Wires and Permalloy Strips at Radiofrequencies

ALVA W. SMITH, J. H. GREGORY,* AND J. T. LYNN**

Mendenhall Laboratory of Physics, The Ohio State University, Columbus, Ohio

(Received September 22, 1945)

The ratio of a.c. to d.c. resistances of iron wires and permalloy strips has been measured in the frequency range 1.5-6.0 megacycles per sec. Empirical equations obtained are compared with existing theoretical equations derived on the assumption of constant permeability.

I. INTRODUCTION

WHEN an electric current flows through a cylindrical conductor in a direction parallel to the axis of the cylinder, magnetic flux lines are set up in circular paths concentric with the axis. If the current is alternating, the time variations of this magnetic flux induce voltages in closed paths in the radial planes of the cylinder. The direction of these induced voltages is such as to decrease the longitudinal current in cylindrical elements of small radii and increase it in elements of larger radii. Thus eddy currents cause the current density to be a maximum at the surface and to decrease as the distance from the surface is increased. As a result of the concentration of current near the surface of the conductor, the a.c. resistance of the conductor is greater than the d.c. resistance. The phenomena here described are usually referred to as skin effect. Since the induced voltages in the radial planes are proportional to the time-rate of change of magnetic flux and the eddy currents produced by these voltages are proportional to the conductivity of

the conductor, skin effect phenomena become more pronounced with increase in frequency, magnetic permeability, and electrical conductivity. An increase in any or all of these factors results in a decrease in the depth of penetration of the current into the conductor and an increase in the ratio of a.c. to d.c. resistance. This statement applies also to conductors of rectangular sections or of other configurations.

Exact formulas for the a.c. to d.c. resistance ratio for straight conductors of circular cross section have been developed by Maxwell, Heaviside, Kelvin, and Rayleigh. By the application of Maxwell's field equations and Ohm's law, this ratio is expressed in terms of a zero-order Bessel function of an argument involving the square root of the product of frequency, permeability, and conductivity.¹ For values of this argument of ten or greater, the solution reduces to the approximate formula

$$R/R_0 = 0.25 + 1.57d(f\mu\sigma/10^9)^{\frac{1}{2}}, \quad (1)$$

in which d is the diameter of the wire in cm; μ the

* Now at Massachusetts Institute of Technology, Cambridge, Massachusetts.

** Now at North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina.

¹ N. W. McLachlan, *Bessel Functions for Engineers* (Oxford Press, New York, 1934), p. 135; August Hund, *High Frequency Measurements* (McGraw-Hill Book Company, Inc., New York, 1933), p. 263; S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio* (John Wiley and Sons, New York, 1944), p. 213.

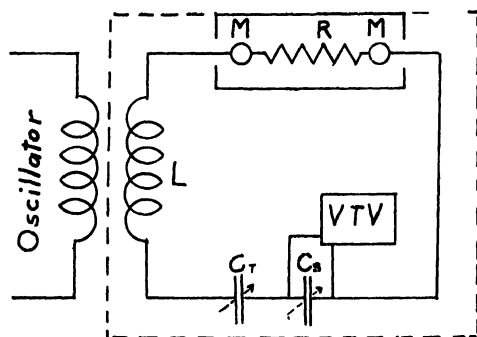


FIG. 1. Plan of the resonant measuring circuit.

magnetic permeability in gauss-oersted⁻¹; and σ the conductivity in mho cm⁻¹.

In the derivation of this formula, the permeability is assumed to be constant. For non-ferromagnetic substances this assumption is valid; for ferromagnetic substances its validity is questionable.

The derivation of a similar formula for conductors of other than circular section has presented considerable mathematical difficulty. Strutt² has obtained a formula for conductors of constant permeability and with cylindrical cross section. His expression is

$$R/R_0 = A + Bf^2, \quad (2)$$

where A and B are constants and f is the frequency. Cockcroft³ has obtained a similar result for conductors of rectangular cross section.

In this paper we are reporting values of the skin-effect resistance ratio for iron wires and permalloy strips in the frequency range 1.5–6.0 megacycles per second.

II. METHOD

The substitution method was used in determining the resistance of each specimen at a given frequency by comparison with the resistances of a set of standards made of No. 28 B.S. constantan wire. Essential features of the apparatus, exclusive of power supplies, oscillator, and General Radio wave meter, type 724-A, are shown in Fig. 1. A Western Electric 305-A tube was used in a Hartley circuit to supply high frequency power through inductive coupling to the

resonant circuit. Tests with the wave meter indicated that the output of the oscillator was essentially free from harmonics. The resonant circuit consisted of an interchangeable inductance L made of quarter-inch copper tubing, the two condensers C_s and C_T , the mercury cups M for receiving the removable resistance R , and the vacuum-tube voltmeter VTV . Fine adjustment of coupling between the oscillator and L was obtained by the use of a worm-drive to rotate L . Each part of the measuring circuit was shielded to prevent unwanted coupling between parts. Tuning was accomplished by means of the standard air condenser C_T , and resonance was indicated by the vacuum-tube voltmeter connected to the terminals of C_s . The circuit of the vacuum-tube voltmeter was of the conventional bridge type in which the vacuum tube acts as one of the bridge arms. In this circuit, a 6 C 6 tube was used as a biased detector.

During a series of measurements at a given frequency, the output of the oscillator, the coupling between it and the resonant circuit, and the setting of C_s were kept constant. The resistance standards and the test samples were connected one at a time to the circuit through the mercury cups to form the resistance R . Condenser C_T was varied until the vacuum-tube voltmeter indicated a maximum current in the circuit. The graph obtained by plotting voltmeter readings against the known resistances of the standards was a straight line. From this graph the values of the resistances of the test samples were obtained by interpolation. A similar procedure was used at each frequency at which measurements were made.

The d.c. resistances of test samples and standards were determined by means of a Mattiessen and Hockin bridge.

Corrections for variations of the resistances of the constantan standards with frequency, amounting to a maximum of 3 percent at 6 megacycles per sec., were applied.

III. DESCRIPTION OF STANDARDS AND TEST SPECIMENS

A set of standard resistances was constructed to cover the range of resistances of the test samples. Suitable lengths of No. 28 B.S. constantan wire were used in making the set. Each

² M. J. O. Strutt, *Electrotechnische Nachrechenstechnik* (1931), Vol. 8, p. 269.

³ J. D. Cockcroft, *Proc. Roy. Soc. [A]*, 122, 533 (1929).

standard was made by soldering each end of the wire to a short section of quarter-inch copper tubing. This resistance element was placed in a straight glass tube and sealed in place at the ends with wax. The protruding ends of the copper tubing were bent to fit into the mercury cups and amalgamated. Lengths of wire in this set ranged from 1.6 to 17 cm.

The test specimens were similarly mounted. The iron wires were cut from chemically pure stock, and the lengths varied from 3.4 to 15.2 cm.

The permalloy samples were cut from a sheet of 78.5 permalloy, having a composition of 78.5 percent nickel and 21.5 percent iron. Samples designated as PA-1 and PA-2 were annealed by heating in a closed electric furnace to 1000°C and cooled with the furnace over an eight-hour period. Samples designated as PQ-1 and PQ-2 were heated to 1000°C, but were withdrawn to cool quickly in the air when the temperature of the furnace had dropped to 600°C. The sample marked PU was not given any heat treatment.

IV. RESULTS AND CONCLUSIONS

Figure 2 shows the a.c. to d.c. resistance ratio as a function of the square root of the frequency in megacycles per sec. for iron wires of the indicated diameters. The empirical equation of these curves is

$$R/R_0 = 0.4 + 1.57d(f\mu\sigma/10^3)^{1/2}, \quad (3)$$

which agrees, within the limits of experimental error, with the theoretical Eq. (1) above. The electrical conductivities were determined from the dimensions and the d.c. resistances of the specimens. The values of magnetic permeabilities were then calculated from the empirical equation. Results are given in Table I.

TABLE I. Data and computed constants for iron wires.

Diam. cm	Length cm	R_0 ohm	σ ohm cm ⁻¹	μ gauss- oersted ⁻¹
0.1019	15.2	0.0204	9.2×10^4	83
0.0823	15.5	0.0311	9.3	77
0.0406	5.0	0.0482	8.0	82
0.0201	3.4	0.0956	11.2	80

The linearity of the curves in Fig. 2 indicates that, with small currents of a few milliamperes, the permeability of each specimen can be regarded as constant within this range of fre-

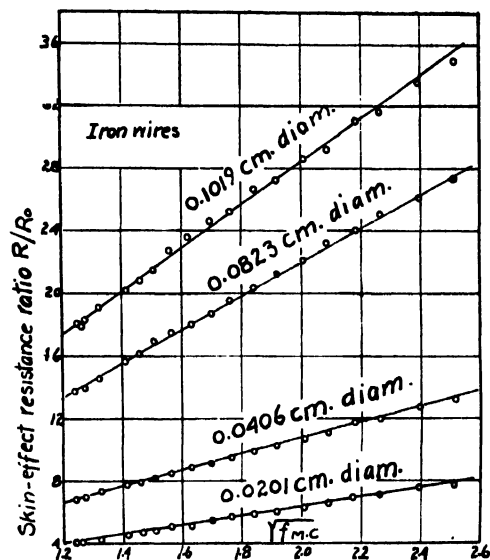


FIG. 2. Ratio of a.c. to d.c. resistance versus the square root of the frequency in megacycles per second.

quencies. Moreover, the deviations of the values of permeability of the individual wires from the mean value were within the limits of experimental error.

The absence of maxima or minima from the curves of Fig. 2 indicates the absence of an anomaly in permeability at 100 meters wavelength reported by Mitiaev.⁴ Our results are in agreement with those of Wait, Brickwedde, and Hall⁵ and those of Strutt and Knoll⁶ in that no such anomalies were found.

In Figs. 3 and 4, the a.c. to d.c. resistance

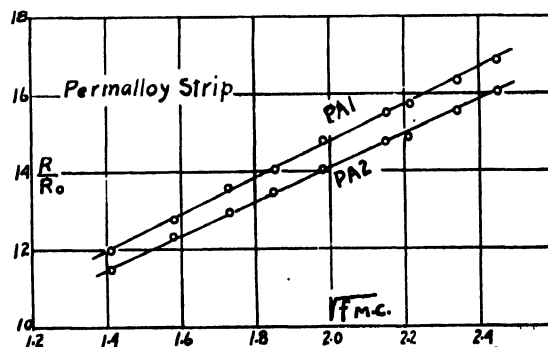


FIG. 3. Ratio of a.c. to d.c. resistance versus the square root of the frequency in megacycles per second. Annealed samples.

⁴ W. K. Mitiaev, *Zeits. f. Physik* **38**, 716 (1926).

⁵ G. R. Wait, F. G. Brickwedde, and E. L. Hall, *Phys. Rev.* **32**, 967 (1928).

⁶ M. J. O. Strutt and K. S. Knoll, *Physica* [7] **2**, 145 (1940).

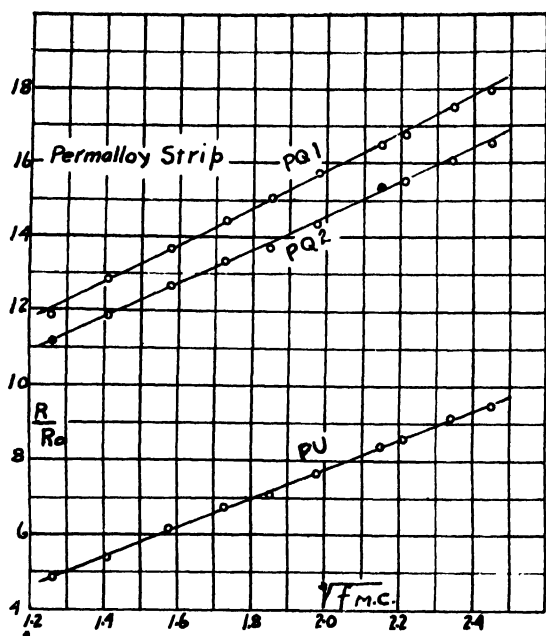


FIG. 4. Ratio of a.c. to d.c. resistance versus the square-root of the frequency in megacycles per second. PQ-1 and PQ-2 quenched samples; PU without heat treatment.

ratios for permalloy strips are plotted as ordinates and the square root of the frequencies, in megacycles per sec., as abscissas. The linearity of the curves indicates that for each specimen the permeability was essentially constant over this frequency range. The general form of equation to fit these data is that of Eq. (2). A more detailed equation in conformity with Cockcroft's theory is

$$R/R_0 = A + f\left(\frac{a}{b}\right) 1.12 \left(\frac{fab\sigma\mu}{10^3}\right)^{\frac{1}{2}}, \quad (4)$$

in which A is an empirical constant for each specimen; f is the frequency in megacycles per sec.; σ is the conductivity in ohm cm^{-1} ; μ is the magnetic permeability; a is the width and b the thickness of the strip in cm; and $f(a/b)$ is a function obtained from Cockcroft's theoretical curves. In attempting to check the agreement between the theoretical values of this function and empirical values for copper calculated from the limited data in the literature, Cockcroft found that the experimental values were a little over half the theoretical values. As a further

TABLE II. Comparison of theoretical and experimental values of $f(a/b)$ for copper strips.

$f^{\frac{1}{2}}$	R/R_0	a	b	σ	$f(a/b)$ exp.	$f(a/b)$ theor.
2.4	1.45	0.15	0.0026	56×10^4		
1.4	1.20				0.48	0.72
2.4	1.525	0.11	0.0026	54		
1.4	1.235				0.66	0.79

check on this point, we made some supplementary measurements on two copper strips and obtained the data and results given in Table II.

The experimental values of $f(a/b)$ from these data are less than the theoretical values, but the disparity is somewhat less than that calculated by Cockcroft. Since the depth of penetration in our copper strip was greater than the half thickness of the strip, it is probable that we cannot come to a definite conclusion as to the reliability of the theoretical values of $f(a/b)$.

However, if we assume the theoretical values of this function to be correct, using Eq. (4), we can calculate from our measurements on permalloy the permeabilities of the strips. Data and computed values are recorded in Table III.

TABLE III. Data and computed constants for permalloy strips.

	a	b	σ	$f(a/b)$ theor.	μ	A
PA-1	0.15	0.040	5.2×10^4	1.52	24	5.4
PA-2	0.11	0.040	4.9	1.62	29	5.2
PQ-1	0.17	0.040	4.9	1.50	27	5.8
PQ-2	0.12	0.040	4.6	1.60	30	5.4
PU	0.10	0.042	4.0	1.66	34	-0.05

If Cockcroft's theoretical values of $f(a/b)$ for an infinite strip are used, the values of μ will be of the order of twice those in the table. Empirical values of the constant A are given in the table. The values of this constant do not show large variations among the heat-treated samples, but these vary widely from that of the untreated sample. No theoretical basis for predicting the value of this constant has been found. The relatively simple form of the empirical equation supports the hope that a definite physical meaning can be assigned to the constant A and Cockcroft's function $f(a/b)$.

Liquid Rise in a Capillary Tube

WESLEY E. BRITTIN

Department of Physics, University of Colorado, Boulder, Colorado

(Received August 20, 1945)

A theory of the dynamics of capillary rise is developed by making certain assumptions as to the nature of the motion of the liquid in the tube. The most important assumptions are that the same forces act on the liquid when it is in an accelerated state of motion as when it is in a steady state, that the surface tension is constant, that the angle of contact between the meniscus of the liquid and the tube wall is constant, and that the wetting of the tube is not a rate-determining factor of the motion. This theory leads to a second-order non-linear differential equation, the solution of which represents the motion of the liquid in the tube. A formal solution of the differential equation is obtained in the form of a double Dirichlet series. Approximations to the series are compared with experimental data, and it is concluded that the agreement between theory and experiment is satisfactory.

INTRODUCTION AND STATEMENT OF THE PROBLEM

C. DECHARME has investigated experimentally and theoretically the motion of a liquid in a capillary tube. The theory¹ of Decharme is somewhat unsatisfactory because of doubtful assumptions and analysis used, and because the final equation of the motion² has three empirical constants which must be evaluated before the equation can be used. An experiment must be performed before the theory can be applied, which defeats at least one purpose of the theory.

W. A. Rense³ has used a stroboscope and motion picture camera to photograph the ascent of water in a capillary tube, and has derived a semi-empirical equation for the motion of the liquid in the tube. G. Pickett⁴ has shown that the theory of Rense is incorrect, and has derived a differential equation for the motion of the liquid which takes into account most of the important forces acting on the liquid. Pickett expressed the solution of his differential equation in terms of an asymptotic relation valid as the time approaches zero, and a finite difference equation valid for the time greater than zero. In the present paper a differential equation for the motion of the liquid in the tube is set up, and a method of solution is discussed. The results of the theory are then compared with experiment. The problem may be stated as follows:

A capillary tube of radius r is touched to the surface of a liquid of density ρ , surface tension σ , and viscosity μ at time $t=0$. Let Z be the extent of rise of the surface of the liquid, measured to the bottom of the meniscus, at time $t \geq 0$, and α the angle that the axis of the tube makes with the horizontal. The problem is then to determine the extent of rise, Z , as a function of the time, t ; $Z=Z(t)$.

THE DIFFERENTIAL EQUATION OF MOTION

An exact hydrodynamical theoretical treatment of the problem would be difficult and has not been attempted. Rather, simplifying assumptions are made in order to obtain an ordinary differential equation for the motion. It is assumed that the same forces act on the liquid when it is in an accelerated state as when it is in a steady state. The amount of liquid above the bottom of the meniscus is neglected which enables one to write $\pi r^2 \rho Z(dZ/dt)$ for the momentum of the liquid in the tube.

The Poiseuille viscous force for streamline flow is $-8\pi\mu Z(dZ/dt)$,⁵ and the component of the weight along the tube is $-\pi r^2 \rho g Z \sin \alpha$, where g is the gravitational force per unit mass. The surface tension force which is responsible for the rise is given by $2\pi r \sigma \cos \vartheta$ where ϑ , the average angle of contact between the surface of the liquid and the wall of the tube, is assumed constant. The forces mentioned above were taken into account by Pickett.

The end of the tube, assumed flat, gives rise to some turbulent flow in the neighborhood of the opening which produces additional drag forces on the liquid. The end-effect drag will be taken to be

¹ C. Decharme, *Ann. Chim. Phys.* **23**, 228-242 (1872); **29**, 415-425, 564-569 (1873); **30**, 145-207, 318-342 (1874).

² C. Decharme, *Ann. Chim. Phys.* **39**, 318-342 (1874).

³ W. A. Rense, *J. App. Phys.* **15**, 436-437 (1944).

⁴ G. Pickett, *J. App. Phys.* **15**, 623 (1944).

⁵ R. von Mises and K. O. Fredricks, *Fluid Dynamics* (Brown University, Providence, Rhode Island, 1941), p. 140.

of magnitude $\frac{1}{4}\pi r^2\rho(dZ/dt)^2$ for the fluid entering the tube.^{6,7} The viscous force brought about by the motion of air in the upper part of the tube is neglected, since it depends upon the density of the air which is small. For extremely long tubes or if the velocity were high, the effect of air could not be neglected. It is further assumed that the rate of wetting of the tube is sufficiently rapid to be insignificant as a factor in the rate of rise of the liquid.

The assumption of streamline flow can be given justification by consideration of the maximum Reynolds number, $Re_{\max}$, for the flow. The Reynolds number, Re , for liquid flow in a tube is defined by

$$Re \equiv (2\rho r/\mu)(dZ/dt),$$

and if this dimensionless ratio is less than a certain critical value, $Re < Re_c \sim 2000$, streamline motion is the rule. The maximum velocity of the liquid is shown below to be

$$(dZ/dt)_{\max} = [(8\sigma \cos \vartheta/5\rho r)]^{\frac{1}{2}},$$

making the maximum Reynolds number

$$Re_{\max} = [(32\rho r\sigma \cos \vartheta/5\mu^2)]^{\frac{1}{2}}.$$

Taking

$$r = 0.05 \text{ cm}, \quad \rho = 1 \text{ g cm}^{-3}, \quad \sigma \cos \vartheta = 75 \text{ dyne cm}^{-1}, \quad \text{and} \quad \mu = 0.005 \text{ poise},$$

$$Re_{\max} \sim 1000 < Re_c,$$

which shows that even for this extreme case the condition for streamline flow is satisfied.

Under the foregoing assumptions Newton's second law of motion yields the following differential equation for the motion of the liquid in the capillary tube:

$$Z\left(\frac{d^2Z}{dt^2}\right) + \frac{5}{4}\left(\frac{dZ}{dt}\right)^2 + bZ\left(\frac{dZ}{dt}\right) + cZ + d = 0, \quad t > 0, \quad (1)$$

where

$$b \equiv \frac{8\mu}{\rho r^2}, \quad c \equiv g \sin \alpha, \quad \text{and} \quad d \equiv -\frac{2\sigma \cos \vartheta}{\rho r}.$$

THE FORMAL SOLUTION OF THE DIFFERENTIAL EQUATION OF MOTION

Assume that the solution of the differential Eq. (1) can be represented by the function series

$$Z = \sum_{i=0}^{\infty} \varphi_i(t)\lambda^i, \quad (2)$$

where λ is an arbitrary parameter independent of t , and the functions $\varphi_i(t)$ are to be determined.⁸ When the series (2) is substituted into Eq. (1) there results

$$\sum_{i=0}^{\infty} \{\alpha_i + \frac{5}{4}\beta_i + b\gamma_i + c\varphi_i\}\lambda^i + d = 0, \quad (3)$$

where

$$\alpha_i \equiv \sum_{k=0}^i \varphi_k \varphi''_{i-k}, \quad \beta_i \equiv \sum_{k=0}^i \varphi_k' \varphi'_{i-k}, \quad \gamma_i \equiv \sum_{k=0}^i \varphi_k \varphi'_{i-k},$$

and primes denote differentiation with respect to the independent variable time. By equating the coefficient of each λ^i to zero, one obtains the set of differential equations for the $\varphi_i(t)$,

$$\varphi_0 \varphi_0'' + \frac{5}{4} \varphi_0'^2 + b \varphi_0 \varphi_0' + c \varphi_0 + d = 0, \quad (4)$$

⁶ W. H. McAdams, *Heat Transfer* (McGraw-Hill Book Company, Inc., New York, 1942), pp. 121-122.

⁷ H. J. Hughes and A. T. Safford, *Hydraulics* (The Macmillan Company, New York, 1911), p. 298.

⁸ The use of similar function series for the solution of non-linear equations is discussed by H. Bateman, *Partial Differential Equations of Mathematical Physics* (Dover Publications, New York, 1944), p. 497.

and

$$\alpha_i + \frac{5}{4}\beta_i + b\gamma_i + c\varphi_i = 0, \quad i > 0. \quad (5)$$

For the present, the case $c=0$ will be excluded, which restricts the motion to non-horizontal flow. For a solution of (4) take

$$\varphi_0 = -d/c, \quad c \neq 0. \quad (6)$$

If the solution (6) is used for φ_0 in (5), it follows that the differential equation for φ_i , $i > 0$, is

$$\varphi_i'' + b\varphi_i' + \frac{c}{\varphi_0}\varphi_i = -\frac{1}{\varphi_0} \sum_{k=1}^{i-1} \left\{ \varphi_k \varphi_{i-k}'' + \frac{5}{4} \varphi_k' \varphi_{i-k}' + b \varphi_k \varphi_{i-k}' \right\}. \quad (7)$$

Put

$$F_i(t) \equiv -\frac{1}{\varphi_0} \sum_{k=1}^{i-1} \left\{ \varphi_k \varphi_{i-k}'' + \frac{5}{4} \varphi_k' \varphi_{i-k}' + b \varphi_k \varphi_{i-k}' \right\}, \quad i > 1. \quad (8)$$

Then if the roots r_1 and r_2 of the quadratic equation

$$\lambda^2 + b\lambda + c/\varphi_0 = 0 \quad (9)$$

are distinct, the general solution of (7) is

$$\varphi_i(t) = \int_0^t G(t-\tau) F_i(\tau) d\tau + C_i \exp(r_1 t) + D_i \exp(r_2 t), \quad (10)$$

where

$$G(t) \equiv \frac{1}{r_1 - r_2} \{ \exp(r_1 t) - \exp(r_2 t) \},$$

and C_i and D_i are arbitrary constants. The solution of the homogeneous part of (7) is the same for all $i > 0$. Therefore without any loss of generality all of the C_i and D_i , $i > 1$, may be set equal to zero,

$$C_i = D_i = 0, \quad i > 1.$$

Putting $\lambda = 1$ in (2) there results

$$Z = \varphi_0 + C_1 \exp(r_1 t) + D_1 \exp(r_2 t) + \sum_{i=2}^{\infty} \int_0^t G(t-\tau) F_i(\tau) d\tau, \quad (11)$$

which, if convergent, represents the general solution of (1). If the roots of Eq. (10) are equal $r_1 = r_2 \equiv r$, then the general solution of (7) is

$$\varphi_i(t) = \int_0^t II(t-\tau) F_i(\tau) d\tau + C_i e^{rt} + D_i t e^{rt}, \quad (12)$$

where $II(t) \equiv t e^{rt}$, and C_i and D_i are arbitrary constants. As before, all C_i and D_i , with the exception of C_1 and D_1 are set equal to zero $C_i = D_i = 0$, $i > 1$. For equal roots then, putting $\lambda = 1$ in (2) results

$$Z = \varphi_0 + C_1 e^{rt} + D_1 t e^{rt} + \sum_{i=2}^{\infty} \int_0^t II(t-\tau) F_i(\tau) d\tau, \quad (13)$$

which, if convergent, is the general solution of (1). It is very unlikely in practice that $r_1 = r_2$, so the solution (11) will be the usual one, rather than (13).

The construction of the series (11) requires the evaluation of the integrals

$$\int_0^t G(t-\tau) F_i(\tau) d\tau$$

which are all elementary, but become more and more cumbersome as i increases. Since the only

terms which appear in the series are of the type $\alpha_{k,l} \exp(kr_1 t) \exp(lr_2 t)$ where $\alpha_{k,l}$ is a constant and k and l are integers, the Dirichlet series

$$Z = \sum_{m,n=0}^{\infty} \alpha_{m,n} \exp[(mr_1 + nr_2)t] \quad (14)$$

can be used to represent the function series (11). To evaluate the coefficients $\alpha_{m,n}$, substitute the series (14) into the differential Eq. (1), whence

$$\sum_{m,n=0}^{\infty} \{\beta_{m,n} + \frac{a}{2}\gamma_{m,n} + b\delta_{m,n} + c\alpha_{m,n}\} \exp[(mr_1 + nr_2)t] + d = 0, \quad (15)$$

where

$$\beta_{m,n} \equiv \sum_{k,l=0}^{m,n} \alpha_{k,l} \alpha_{m-k,n-l} (kr_1 + lr_2)^2, \quad \gamma_{m,n} \equiv \sum_{k,l=0}^{m,n} \alpha_{k,l} \alpha_{m-k,n-l} (kr_1 + lr_2) \{(m-k)r_1 + (n-l)r_2\},$$

and

$$\delta_{m,n} \equiv \sum_{k,l=0}^{m,n} \alpha_{k,l} \alpha_{m-k,n-l} (kr_1 + lr_2).$$

Equation (1) will be satisfied formally if

$$\beta_{0,0} + \frac{a}{2}\gamma_{0,0} + b\delta_{0,0} + c\alpha_{0,0} + d = 0, \quad (16)$$

and

$$\beta_{m,n} + \frac{a}{2}\gamma_{m,n} + b\delta_{m,n} + c\alpha_{m,n} = 0, \quad \binom{m}{n} \neq \binom{0}{0}. \quad (17)$$

Therefore

$$\alpha_{0,0} = -d/c, \quad c \neq 0, \quad (18)$$

$$\alpha_{1,0} \left(r_1^2 + br_1 + \frac{c}{\alpha_{0,0}} \right) = 0, \quad (19)$$

$$\alpha_{0,1} \left(r_2^2 + br_2 + \frac{c}{\alpha_{0,0}} \right) = 0, \quad (20)$$

and

$$\alpha_{m,n} = \frac{-1}{\alpha_{0,0} \{(mr_1 + nr_2)^2 + b(mr_1 + nr_2) + c/\alpha_{0,0}\}}$$

$$\times \sum'_{k,l=0}^{m,n} \alpha_{k,l} \alpha_{m-k,n-l} \left\{ -\frac{1}{2}(kr_1 + lr_2)^2 + (kr_1 + lr_2) \left[\frac{a}{2}(mr_1 + nr_2) + b \right] \right\}, \quad \binom{m}{n} \neq \begin{cases} \binom{0}{0} \\ \binom{0}{1} \\ \binom{1}{0} \end{cases}, \quad (21)$$

where $\sum'$ denotes summation omitting the term $\binom{k}{l} = \binom{m}{n}$. If $\alpha_{1,0} \neq 0$, and $\alpha_{0,1} \neq 0$, r_1 and r_2 must be the roots of the quadratic

$$x^2 + bx + c/\alpha_{0,0} = 0, \quad (22)$$

which was to be expected according to Eq. (9), since by (6) and (18) $\alpha_{0,0} = \varphi_0$. If the Dirichlet series (14) is convergent, it represents the general solution of (1) for the case of distinct roots, $r_1 \neq r_2$. The

recurrence relation (21) gives all of the coefficients $\alpha_{m,n}$, in terms of the two arbitrary and independent constants $\alpha_{0,1}$, and $\alpha_{1,0}$.

The evaluation of the constants $\alpha_{0,1}$ and $\alpha_{1,0}$ is made possible by means of the initial conditions

$$Z(0) = 0, \quad (dZ/dt)_0 = V_0. \quad (23)$$

The initial velocity V_0 is assumed to be finite as it must be physically, and thus is obtained from the differential equation by setting $Z=0$,

$$V_0 = (-\frac{1}{4}d)^{\frac{1}{2}}. \quad (24)$$

In an actual problem one must use an approximation to the series (14), and the character of the roots r_1 and r_2 determine to a great extent the number of terms required for a good approximation. The roots r_1 and r_2 will be real and negative if $b^2 - 4c/\alpha_{0,0} > 0$, and a simple calculation shows that for most liquids r_1 and r_2 will be real and negative if the capillary tube has a radius in the neighborhood of 0.010 cm or less. The following treatment applies to real roots, and r_1 will be taken to be of smaller magnitude than r_2 , $|r_1| < |r_2|$. If $|r_2|$ is large compared to $|r_1|$, a good approximation may be had by using only one term in r_2 and n terms in r_1 ,

$$Z \sim \alpha_{0,0} + \alpha_{0,1} \exp(r_2 t) + \sum_{k=1}^n \alpha_{k,0} \exp(k r_1 t), \quad (25)$$

where n is restricted by $n \leq |r_2/r_1|$. By Eq. (21), the coefficient $\alpha_{k,0}$ has a factor $(\alpha_{1,0})^k$, so for convenience define α_k by

$$\alpha_k \equiv \alpha_{k,0} / (\alpha_{1,0})^k \quad (26)$$

Equation (25) then becomes

$$Z \sim \alpha_{0,0} + \alpha_{0,1} \exp(r_2 t) + \sum_{k=1}^n \alpha_k (\alpha_{1,0})^k \exp(k r_1 t). \quad (27)$$

The evaluation of $\alpha_{0,1}$ and $\alpha_{1,0}$ is made by utilizing relations (23) and (24), from which it is evident that $\alpha_{0,1}$ and $\alpha_{1,0}$ must satisfy

$$\alpha_{0,0} + \alpha_{0,1} + \sum_{k=1}^n \alpha_k (\alpha_{1,0})^k = 0, \quad (28)$$

and

$$r_2 \alpha_{0,1} + \sum_{k=1}^n r_1 k \alpha_k (\alpha_{1,0})^k = V_0. \quad (29)$$

The elimination of $\alpha_{0,1}$ between (28) and (29) gives rise to an algebraic equation of the n th degree in $\alpha_{1,0}$,

$$V_0 + r_2 \alpha_{0,0} = \sum_{k=1}^n (k r_1 - r_2) \alpha_k (\alpha_{1,0})^k. \quad (30)$$

The real root of Eq. (30) having the smallest magnitude will be taken as $\alpha_{1,0}$, while the value of $\alpha_{0,1}$ will be obtained directly from (28). The α_k are given by

$$\alpha_k \equiv \frac{-1}{\alpha_{0,0}(k^2 r_1^2 + k r_1 + c/\alpha_{0,0})} \sum_{i=0}^{k-1} \alpha_i \alpha_{k-i} r_1 i (-\frac{1}{2} r_1 i + \frac{1}{2} k r_1 + b), \quad k > 1. \quad (31)$$

Equations (28)–(30) will be subsequently applied to actual physical data.

It is interesting to note that the transformation $(dZ/dt) = 1/W$ reduces the original differential Eq. (1) to the first-order equation

$$\left(\frac{dW}{dZ}\right) = -\frac{5}{4} \frac{W}{Z} + b W^2 + c W^3 + d \frac{W^3}{Z}. \quad (32)$$

The solution of (32) can be used to find t in terms of Z ,

$$t = \int W(Z) dZ. \quad (33)$$

COMPARISON OF THEORY WITH EXPERIMENT

If the series (14) is to have any meaning as the solution of the differential equation, it must necessarily converge. The mathematical analysis of the convergence or divergence of the series is encumbered by the fact that the coefficients, $\alpha_{m,n}$, are not known as explicit functions of m and n , but are known only through the multiple-term recurrence relation (21). The difficulty of applying standard convergence tests to the double series (14) is thus apparent.

An indication, but certainly not a proof, of the convergence of the series (14) is obtained by taking an approximation to it using actual physical data, and then comparing the result with experiment. As one example, the data of Rense³ are used, and the temperature is assumed to be 26°C. A five-term approximation from Eq. (25) is

$$Z \sim \alpha_{0,0} + \alpha_{0,1} \exp(r_1 t) + \alpha_{1,0} \exp(r_1 t) - \frac{\alpha_{1,0}^2}{\alpha_{0,0}} \left\{ \frac{r_1^2 + \frac{5}{4} r_1^2 + b r_1}{4 r_1^2 + 2 b r_1 + c / \alpha_{0,0}} \right\} \exp(2 r_1 t) \\ + \frac{\alpha_{1,0}^3}{\alpha_{0,0}^2} \left\{ \frac{r_1^2 + \frac{5}{4} r_1^2 + b r_1}{4 r_1^2 + 2 b r_1 + c / \alpha_{0,0}} \right\} \left\{ \frac{5 r_1^2 + 5 r_1^2 + 3 b r_1}{9 r_1^2 + 3 b r_1 + c / \alpha_{0,0}} \right\} \exp(3 r_1 t). \quad (34)$$

Using relations (22)–(24), Eq. (34) becomes

$$Z \sim 4.633 - 0.585 e^{-66.84 t} - 2.187 e^{-3.165 t} - 1.037 e^{-0.33 t} - 0.787 e^{-0.495 t} \quad (35)$$

which is plotted in Fig. 1.

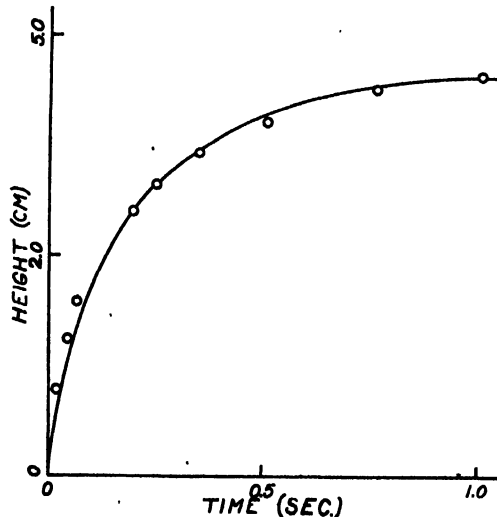


FIG. 1. Rise of water in a vertical tube. Circles represent data of Rense. Solid curve represents Eq. (36).

A further comparison is obtained from the results of Decharme⁹ on water rising in a capillary tube inclined at an angle of 45°. The five-term approximation given by Eq. (25) is

$$Z \sim 11.745 - 0.0682 e^{-255.4 t} - 6.108 e^{-0.33 t} - 3.137 e^{-0.46 t} - 2.432 e^{-0.69 t}. \quad (36)$$

⁹C. Decharme, Ann. Chim. Phys. 30, 162 (1874).

TABLE I. Rise of pure water in a capillary tube of radius 0.01777 cm, temperature of water 19.5°C, and angle of inclination of the tube 45°.

Time (sec.)	Z (cm) Eq. (36)	Z (cm) Eq. (37)	Z (cm) Eq. (38)	Z (cm) Eq. (39)	Z (cm) Decharme	Time (sec.)	Z (cm) Eq. (36)	Z (cm) Eq. (37)	Z (cm) Eq. (38)	Z (cm) Eq. (39)	Z (cm) Decharme
0	0.00	0.00	0.00	0.00	0.00	8				10.91	10.87
1	3.69	4.19	4.61	4.96	6.20	9				11.09	11.05
2	6.03	6.46	7.00	7.22	7.87	10				11.23	11.20
3	7.59	7.96	8.34	8.52	8.80	12				11.43	11.40
4	8.66	8.96	9.24	9.36	9.48	14				11.55	11.54
5	9.42	9.57	9.84	9.94	9.98	16				11.62	11.62
6	9.97	—	—	10.36	10.36	18				11.67	11.66
7				10.67	10.65	20				11.70	11.70

The six-term approximation is

$$Z \sim 11.745 - 0.085e^{-258.4t} - 5.591e^{-0.23t} - 2.629e^{-0.46t} - 1.866e^{-0.69t} - 1.574e^{-0.92t}. \quad (37)$$

The eight-term approximation is

$$Z \sim 11.745 - 0.077e^{-258.4t} - 5.123e^{-0.23t} - 2.207e^{-0.46t} - 1.435e^{-0.69t} - 1.110e^{-0.92t} - 0.944e^{-1.15t} - 0.849e^{-1.38t}, \quad (38)$$

and finally the ten-term approximation is

$$Z \sim 11.745 - 0.100e^{-258.4t} - 4.906e^{-0.23t} - 2.024e^{-0.46t} - 1.260e^{-0.69t} - 0.933e^{-0.92t} - 0.761e^{-1.15t} - 0.655e^{-1.38t} - 0.561e^{-1.61t} - 0.545e^{-1.84t}. \quad (39)$$

Table I gives the values of Z as a function of time for the approximations (36)–(39), and for the experimental values obtained by Decharme. It is to be noted that the agreement of the approximations with the experimental values does improve with the number of terms in the approximation.

HORIZONTAL FLOW

For horizontal flow, $\alpha = 0$, and Eq. (1) becomes

$$ZZ'' + \frac{1}{4}(Z')^2 + bZZ' + d = 0, \quad (40)$$

or

$$(ZZ')' + \frac{1}{4}(Z')^2 + bZZ' + d = 0. \quad (41)$$

If the tube is of small bore, $\frac{1}{4}(Z')^2$ can be neglected, which means that the end-effect drag is neglected, and Eq. (41) is linear in (ZZ') . The solution, neglecting $\frac{1}{4}(Z')^2$, is

$$ZZ' = -de^{-bt} \int e^{bt} dt = -\frac{d}{b} \{1 + C_1 e^{-bt}\}, \quad (42)$$

where C_1 is a constant of integration. When $t = 0$, $(ZZ') = 0$, and $C_1 = -1$. Integrating (42) one obtains

$$Z^2 = -\frac{2d}{b^2} \{bt + e^{-bt} + C_2\}, \quad (43)$$

where C_2 is a constant of integration. When $t = 0$, $Z = 0$, and $C_2 = -1$, which when substituted into (43) yields

$$Z^2 = -\frac{2d}{b^2} \{bt + e^{-bt} - 1\}. \quad (44)$$

Substituting $b = 8\mu/\rho r^2$ and $d = -2\sigma \cos \theta/\rho r$ into (44), one obtains the approximate solution for

horizontal flow,

$$Z^2 = \frac{\rho \sigma r^3 \cos \beta}{16\mu^2} \left\{ \left(\frac{8\mu}{\rho r^2} \right) t + \exp \left[- \left(\frac{8\mu}{\rho r^2} \right) t \right] - 1 \right\}. \quad (45)$$

If liquid were forced through a capillary tube under pressure, a constant term would be added to d , and if conditions for streamline flow were not violated, the above theory should still be satisfactory. For large values of t , Z^2 is asymptotically linear in t according to Eq. (45), which is in qualitative, if not quantitative, agreement with experiment.^{10,11}

¹⁰ J. M. Bell and F. K. Cameron, *J. Phys. Chem.* **10**, 658-674 (1906).

¹¹ R. L. Peek and D. A. McLean, *Ind. Eng. Chem. Anal. Ed.* **6**, 85-90 (1934).

Shunt and Series Sections of Transmission Line for Impedance Matching

C. T. TAI

Cruft Laboratory, Harvard University, Cambridge, Massachusetts

(Received August 20, 1945)

The problem of using shunt and series sections of transmission line as matching devices is here investigated mathematically. By making use of hyperbolic functions the condition of match for both cases can be expressed in a very simple way. Both devices when used alone have a limited range with regard to the impedance to be matched. The exact range within which matching is possible is shown graphically. By using one additional matching network each device can be designed to match impedance of any value. The complete systems still preserve their simplicity in structure.

INTRODUCTION

PARALLEL lines or coaxial lines are often used to transmit power from a generator to a load. In practice, it is desirable to eliminate standing waves on these lines, that is, to make the lines non-resonant. Under this condition the efficiency of transmission along the line is a maximum. A line is non-resonant if it is terminated in an impedance equal to the characteristic impedance of the line. A load impedance not equal to the characteristic impedance of the line can be made to match the line by means of a matching device. This is illustrated in Fig. 1, where Z_L is the load

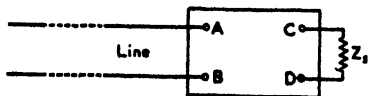


FIG. 1. Matching section for a transmission line.

impedance and $ABCD$ is a matching device. By a proper design of this device the input impedance looking from AB toward the load can be made equal to the characteristic impedance of the line. It is known that shunt sections and series sections when properly designed can be used as matching devices.^{1,2} However, little has been published

about the limitations of these two devices. The present paper discusses them in detail in two separate parts.

PART I

Shunt Sections of Line

General Equations and Their Solutions

The structure of shunt sections of line is shown in Fig. 2. It consists of two sections of line of the same characteristic impedance R_c as the main line, and of length s_1 and s_2 . Physically, either branch may be considered as an extension of the main line, with the other branch connected in shunt. It is assumed in the following analysis that radiation from all sections of line is negligible, and that the coupling between lines I and II is so loose that mutual impedance may be neglected. These two assumptions are legitimate if the separation of the wires forming each line is small

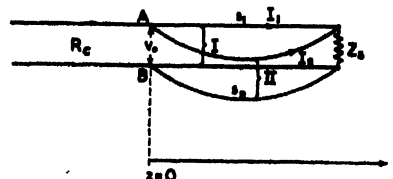


FIG. 2. Shunt section of transmission line for impedance matching.

¹ Ronald King, "Transmission line theory and its application," *E. App. Phys.* **14**, 577 (November, 1943).

² Andrew Alford, "High frequency transmission-line network," *Comm.* **17**, No. 3, 301 (January, 1939).

compared with its length and if the two branches are not very close. Then the current and voltage of either of them, expressed in complex form, satisfy the following two differential equations:

$$dI/ds = -YV, \quad (1)$$

$$dV/ds = -ZI, \quad (2)$$

where Z and Y are complex line constants.

The solution of these two equations in terms of hyperbolic functions is

$$I = C_1 \cosh ks + C_2 \sinh ks, \quad (3)$$

$$V = -Z_c(C_1 \sinh ks + C_2 \cosh ks), \quad (4)$$

where k is the complex propagation constant and Z_c the characteristic impedance of the lines. They

are related to Z and Y as follows:

$$k = \alpha + j\beta = (ZY)^{1/2}, \quad (5)$$

$$Z_c = R_c(1 - j\phi_c) = (Z/Y)^{1/2}. \quad (6)$$

It is assumed in the following analysis that attenuation along the two sections of lines is negligible so that the complex constant k is purely imaginary, being equal to $j\beta$, and Z_c is purely resistive, being equal to R_c .

If one defines the complex terminal functions,¹ $\theta = \rho + j\Phi$, for the terminal impedance Z_s , by setting

$$Z_s = R_c \coth \theta, \quad (7)$$

the input impedance looking from AB in Fig. 2 can be found to be

$$Z_{in} = R_c \frac{\cosh^2 \theta \cos \beta s_1 \cos \beta s_2 - \cosh(\theta + j\beta s_1) \cosh(\theta + j\beta s_2)}{\sinh 2\theta - \sinh(2\theta + j\beta(s_1 + s_2)) + j \cosh^2 \theta \sin \beta(s_1 + s_2)}. \quad (8)$$

The condition of match requires that Z_{in}/R_c be equal to unity. Setting Z_{in}/R_c in (8) equal to unity and separating the real and imaginary parts one obtains the following solutions for s_1 and s_2 :

$$\tan \beta \left(\frac{s_1 + s_2}{2} \right) = \frac{\cos 2\Phi - e^{-2\rho}}{2 \sin 2\Phi}, \quad (9)$$

$$\tan \beta \left(\frac{s_1 - s_2}{2} \right) = \left[\frac{(5e^{-2\rho} + 3 \cos 2\Phi)(e^{-2\rho} - \cos 2\Phi)}{4(1 - e^{-4\rho})} \right]^{1/2}. \quad (10)$$

These two equations determine completely the length of the two sections for different values of ρ and Φ or different values of Z_s , provided the solutions for s_1 and s_2 are real and positive.

It is to be noted that there is no practical limitation upon the value of ρ and Φ in order to yield a real value of $s_1 + s_2$, since the right-side expression of (9) is always real. But in order that the value of $s_1 - s_2$ be also real, ρ and Φ have to satisfy the following condition,

$$\frac{(5e^{-2\rho} + 3 \cos 2\Phi)(e^{-2\rho} - \cos 2\Phi)}{1 - e^{-4\rho}} \geq 0. \quad (11)$$

As results from (7), for each given pair of values of R_s and X_s there is a corresponding pair of values of ρ and Φ , and *vice versa*. The inequality (11), therefore, implies that matching is possible

only for those impedances of which the corresponding values of ρ and Φ satisfy this inequality.

In order to carry on the discussion in terms of the more familiar parameters R_s and X_s jointly, it is necessary to review briefly some of the analytical relations between ρ and Φ with R_s and X_s (for a fuller discussion the reader is referred to reference 1). The terminal functions ρ and Φ defined by (7) can be directly expressed in terms of R_s/R_c and X_s/R_c . They are

$$\rho = \frac{1}{2} \tanh^{-1} \left[\frac{2r_s}{r_s^2 + x_s^2 + 1} \right], \quad (12)$$

$$\Phi = \frac{1}{2} \tan^{-1} \left[\frac{-2x_s}{r_s^2 + x_s^2 - 1} \right], \quad (13)$$

where r_s and x_s are defined for a low-loss line by

$$r_s = \frac{R_s}{R_c}, \quad x_s = \frac{X_s}{R_c}. \quad (14)$$

The values of ρ and Φ may also be determined graphically with sufficient accuracy from a "circle diagram" as discussed in reference 1. This diagram is reproduced in Fig. 3.

The condition of match as expressed by (11) can now be used to find the region in the r_s - x_s plane within which matching is possible. Substituting the values of ρ and Φ given by (12) and (13) into the left side of (11) and setting the resulting expression equal to zero, one obtains the

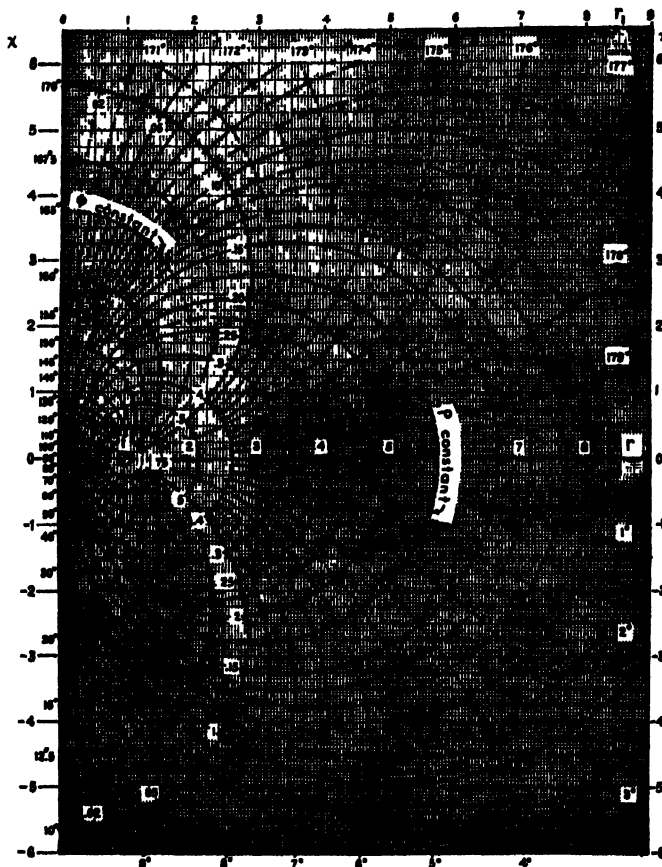


FIG. 3. Circle diagram (from reference 1).

following two equations:

$$r_s = 1, \quad (15)$$

$$4(r_s^2 + x_s^2) - 5r_s + 1 = 0. \quad (16)$$

Equation (15) is the equation of a vertical straight line passing through $r_s = 1$, and (16) is that of a circle with center at $(\frac{5}{8}, 0)$ and of radius equal to $\frac{3}{8}$. These two curves are plotted in Fig. 4. They define two distinct zones in the plane of $r_s - x_s$. Matching is possible only for those values if r_s and x_s lie to the left of the line $r_s = 1$, with the region enclosed by the circle C_{sh} excluded. On the other hand, matching is not possible by using shunt sections alone if the values of r_s and x_s lie to the right of the line or inside the circle C_{sh} .

Shunt Sections with Additional Section of Line

The above analysis deals with shunt sections only and the restrictions thus found are inherent to this type of network. It is obvious, however, that if a simple additional impedance-transforming network be added such that any impedance can be transformed into a value which satisfies the matching condition of the shunt

sections, then the limitation no longer exists. The simplest way of doing this is to insert an additional section of line of the same characteristic impedance between the shunt sections and the load, as shown in Fig. 5. It is to be added here that an analysis of this system applying the theorem of conjugate impedance is given in reference 2, where the system is called "re-entrant



FIG. 4. Loci of match for shunt sections.

lines." But because of the close similarity between the shunt sections and the series section, which will appear in Part II, a more specific term is used in this paper.

Referring to reference 1 it is seen that the values of ρ and Φ corresponding to the impedance looking from CD in Fig. 5 to the right are

$$\begin{cases} \rho_{in} = \rho + \alpha s_3, \\ \Phi_{in} = \Phi + \beta s_3, \end{cases} \quad (17)$$

where s_3 denotes the length of Section III. Since α is assumed to be small, (17) can practically be written as

$$\begin{cases} \rho_{in} = \rho, \\ \Phi_{in} = \Phi + \beta s_3. \end{cases} \quad (18)$$

The only modification of ρ and Φ due to the presence of Section III is therefore a change in Φ only.

Referring back to Figs. 3 and 4, suppose a pair of values of ρ and Φ to be given that corresponds to a point lying outside the region where matching is possible, say, $\rho = 0.5$ and $\Phi = 15^\circ$. Ordinarily, that is, without the insertion of Section III, matching would be impossible. By inserting an additional section of line of electrical length, say, 30° , a shift is made to a new point where $\rho = 0.5$ and $\Phi = 45^\circ$. This new point is well within the domain where matching is possible. It is to be noted that as far as matching is concerned the exact value of s_3 is not critical. But with s_3 so fixed there is only one pair of values s_1 and s_2 that makes the main line non-resonant. Since all the constant- ρ circles have some part of them included in the matching region, it is always possible to shift any given pair of values of ρ and Φ into this region by the insertion of an additional section of line. Any given load Z_L can in this way be matched to a given line.

Graphical Representation of Solutions

For practical purposes, the values of s_1/λ and s_2/λ can be solved for in terms of ρ and Φ from (9) and (10). They are

$$s_1/\lambda = \frac{1}{2\pi} [\tan^{-1} A + \tan^{-1} B], \quad (19a)$$

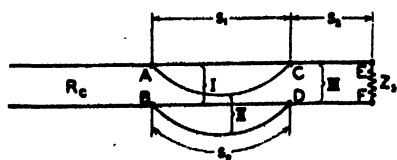


FIG. 5. Shunt sections with additional section of line.

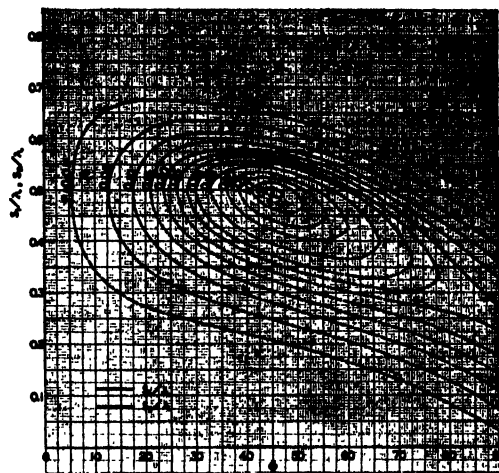


FIG. 6. s_1/λ and s_2/λ vs. ρ and Φ .

$$s_2/\lambda = \frac{1}{2\pi} [\tan^{-1} A - \tan^{-1} B], \quad (19b)$$

where A and B are defined by the following equations:

$$A = \frac{\cos 2\Phi - e^{-2\rho}}{2 \sin 2\Phi}; \quad (20a)$$

$$B = \left[\frac{(5e^{-2\rho} + 3 \cos 2\Phi)(e^{-2\rho} - \cos 2\Phi)}{4(1 - e^{-4\rho})} \right]^{1/2}. \quad (20b)$$

The values of s_1/λ and s_2/λ are plotted in Fig. 6 for different values of ρ and Φ . The mirror part corresponding to $90^\circ < \Phi < 180^\circ$ is omitted. When ρ and Φ are known s_1 and s_2 can be readily determined from these curves. Numerical examples concerning the design of shunt sections will be given at the end of the paper.

PART II

Series Section of Line

General Solution

It is well known that lines of different characteristic impedance may also be used for matching. This arrangement is usually called a series transformer. From the point of view of its structure and also as a counterpart of the shunt sections described above, the name "series section" will be used. Its structure is shown diagrammatically in Fig. 7, where R_L' and R_0 denote, respectively, the characteristic impedance of the line to be matched and that of the line used for matching. It has been derived in reference 1 that if the

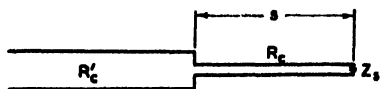


FIG. 7. Series section of transmission line for impedance matching.

attenuation along the line R_c is negligible, the general condition for match is

$$R_c' > R_c \begin{cases} R_c' = R_c \coth \rho, \\ \beta s = n\pi - \Phi, \end{cases} \quad (21)$$

$$\beta s = n\pi - \Phi, \quad (22)$$

$$R_c' < R_c \begin{cases} R_c' = R_c \tanh \rho, \\ \beta s = \left(\frac{2n+1}{2}\right)\pi - \Phi, \end{cases} \quad (23)$$

$$\beta s = \left(\frac{2n+1}{2}\right)\pi - \Phi, \quad (24)$$

where ρ and Φ denote the terminal functions of the load impedance defined in terms of R_s , X_s , and the characteristic impedance R_c which is the unknown. They are given by (12) and (13). In case of resistive termination both (21–22) and (23–24) with $n=1$ and 0, respectively, reduce to

$$\begin{cases} R_c = (R_c' R_s)^{1/2}, \\ \beta s = \frac{\pi}{2} \end{cases} \quad (25)$$

$$\beta s = \frac{\pi}{2} \quad (26)$$

This shows that any purely resistive load can be matched theoretically by choosing a proper value of R_c for the matching line of length $\lambda/4$. A more general case will now be treated in which the load impedance includes reactance as well as resistance. The question is whether it is always possible for any given values of Z_s to find values for R_c and s such that the main line with characteristic impedance R_c' can be made non-resonant.

The explicit solution for R_c in (21) or (23) is

$$R_c = \left[\frac{R_c' (R_s^2 + X_s^2) - R_s R_c'}{R_s - R_c'} \right]^{1/2}. \quad (27)$$

In order for R_c to be real, the following conditions must be satisfied

$$R_s > R_c', \quad R_s^2 + X_s^2 > R_s R_c', \quad (28)$$

$$R_s < R_c', \quad R_s^2 + X_s^2 < R_s R_c'. \quad (29)$$

Condition (28) is always true, while (29) shows that the values of R_s and X_s must lie within certain definite domains for a given value of R_c' . As in the discussion of shunt sections, it is possible to find the regions in the $r_s' - x_s'$ plane within which matching is possible. These regions

are bounded by the following curves:

$$r_s' = 1, \quad (30)$$

$$r_s'^2 + x_s'^2 - r_s' = 0, \quad (31)$$

where r_s' and x_s' denote, respectively, R_s/R_c' and X_s/R_c' . Equation (30) is the equation of a vertical straight line passing through $r_s' = 1$ and (31) is that of a circle with center at $(\frac{1}{2}, 0)$ and of radius equal to $\frac{1}{2}$. They are plotted in Fig. 8. As in the shunt case, the curves defined again two distinct zones. Matching is possible only for those values of r_s' and x_s' that lie to the right of the line $r_s' = 1$, or within the circle C_s . On the other hand, matching is not possible, using series section alone, if the values of r_s' and x_s' lie outside the circle and to the left of the line $r_s' = 1$.

The condition of match for the series section, insofar as R_c is concerned, has been discussed in terms of r_s' and x_s' ; the solution for R_c , (27), is also expressed in terms of R_s , X_s , and R_c' . The solutions for s , (22) and (24), however, are still expressed in terms of Φ which is itself a function of R_s , X_s , and R_c . To avoid confusion, it is desirable to obtain an expression for s in terms of R_s , X_s , and R_c' directly. By making use of formula (13), one obtains the following expression of Φ ,

$$\begin{aligned} \Phi &= \frac{1}{2} \tan^{-1} \left[\frac{-2(X_s/R_c)}{(R_s/R_c)^2 + (X_s/R_c)^2 - 1} \right] \\ &= \frac{1}{2} \tan^{-1} \left[\frac{-2mx_s'}{r_s'^2 + x_s'^2 - m^2} \right], \end{aligned} \quad (32)$$

where m denotes the ratio R_c/R_c' which is given by (27), i.e.,

$$m = \frac{R_c}{R_c'} = \left[\frac{r_s'^2 + x_s'^2 - r_s'}{r_s' - 1} \right]^{1/2}. \quad (33)$$

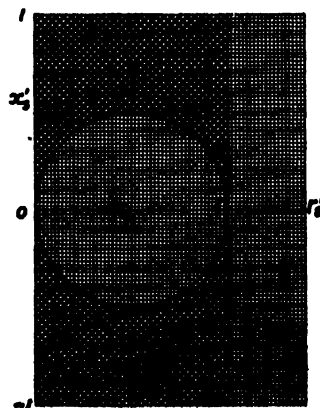


FIG. 8. Loci of match for series section.

The complete solution for R_e and s can then be written as follows,

$$R_e = mR_e' = R_e' \left(\frac{r_s'^2 + x_s'^2 - r_s'}{r_s' - 1} \right)^{\frac{1}{2}} \quad (34)$$

$$m < 1 \quad \left\{ \beta s = n\pi - \frac{1}{2} \tan^{-1} \left[\frac{-2mx_s'}{r_s'^2 + x_s'^2 - m^2} \right] = n\pi - \Phi, \quad (35)$$

$$m > 1 \quad \left\{ \beta s = \frac{(2n+1)}{2} \pi - \frac{1}{2} \tan^{-1} \left[\frac{-2mx_s'}{r_s'^2 + x_s'^2 - m^2} \right] = \left(\frac{2n+1}{2} \right) \pi - \Phi. \quad (36)$$

The form of solution for s in terms of Φ has been preserved, for it is more convenient as will be apparent in the illustrative example given later.

Series Section with Additional Section of Line

As in the case of shunt sections, the series section also is limited in its application if it is used alone. But here, as before, it is possible to shift the values of ρ' and Φ' (where ρ' and Φ' are the equivalent of r_s' and x_s' so they are defined in terms of R_e , X_e , and R_e' , not R_e), practically of Φ' only, by connecting an additional section of line between the series section and the load. The complete system then has the structure shown in Fig. 9. It is evident from Fig. 8 and Fig. 3 that the length of this additional section of line need not be longer than $\lambda/4$.

Graphical Solution

Instead of plotting the values of R_e and s against r_s' and x_s' , it is simpler in this case to use a graphical construction. If one treats m as a parameter, (33) defines a family of circles which are plotted in Fig. 10. Besides the point (1, 0) each circle crosses the r_s' -axis at $(m^2, 0)$. This diagram enables one to determine R_e/R_e' quite rapidly. If one enters the diagram at a point where there is no available circle; instead of using interpolation, one can determine m by constructing the line AB as shown in Fig. 10 and another line BC perpendicular to it. The square root of OC then gives the value of m . Once m has been determined, one can make use of the circle

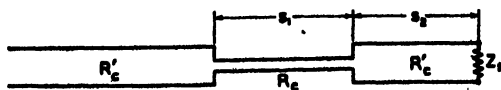


FIG. 9. Series section with additional section of line.

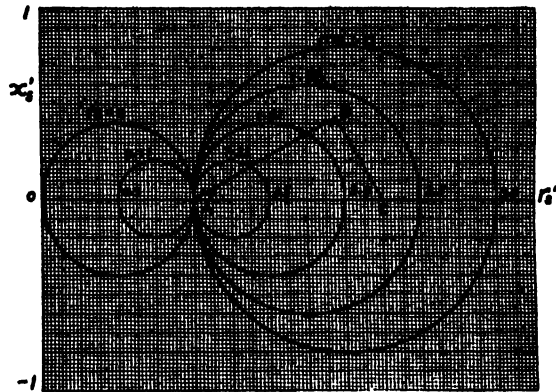


FIG. 10. Constant m circles.

diagram to determine the value of Φ , and hence the value of s by (35) or (36).

Numerical Examples

Four examples illustrating the practical design of typical shunt and series sections are given below.

Shunt sections:

Case 1.—Shunt section alone, i.e., terminated directly in the load.

Suppose $R_e = 400$ ohms, $Z_L = 200 - j400$ ohms. This corresponds to $r_s = 0.5$ and $x_s = -1$. Referring to Fig. 4, this point lies inside the matching region. From the circle diagram one finds that the values of ρ and Φ corresponding to the given values of r_s and x_s are 0.24 nepers and 41.5° , respectively. From Fig. 6 one finds $s_1/\lambda = 0.572$ and $s_2/\lambda = 0.350$.

Case 2.—Shunt sections with additional section of line.

Suppose $R_e = 400$ ohms, $Z_L = 800 + j0$ ohms. This corresponds to $\rho = 0.55$ and $\Phi = 0$. It is seen from Fig. 4 that this pair of values of r_s and x_s lies outside the matching region. Let an additional section of line of length $s_2 = \lambda/8$, or $\beta s_2 = 45^\circ$ be

inserted. Then the effective value of Φ changes to 45° while ρ remains unchanged. This new pair of values of ρ and Φ now lies well inside the matching region. From Fig. 6 the values of s_1/λ and s_2/λ are found to be 0.535 and 0.414, respectively.

Series sections:

Case 3.—Series section alone.

Suppose $R_c' = 400$ ohms, and $Z_s = 800 - j40$ ohms. This corresponds to $r_s' = 2$ and $x_s' = -0.1$. Referring to Fig. 8, this point lies inside the matching region. The value of $m = R_c/R_c'$ determined from Fig. 10 is 1.42; so $R_c = 568$ ohms. With this value of R_c one finds $r_s = 1.41$, $x_s = -0.0705$. The corresponding value of Φ as determined from the circle diagram is 5° . Using (36), with $n = 0$, one obtains $s/\lambda = 0.236$.

Case 4.—Series section with additional section of line

Suppose $R_c' = 400$ ohms, and $Z_s = 200 + j500$ ohms. The given point lies outside the matching region in Fig. 8. The corresponding values of ρ' and Φ' are 0.186 nepers and 144° . Suppose Φ' is increased by 36° by inserting an additional length section of line of length s_2 equal to 0.1λ . The effective Φ' then shifts to 180° while ρ' remains unchanged. From Fig. 10 one finds $m = 2.33$ for this new point, hence $R_c = 930$ ohms. The length of the section s_1 is $\lambda/4$.

CONCLUSION

The preceding analysis shows that a close similarity exists between the use of shunt and series sections of line. When used alone, both are limited to a certain range of impedances. But there is a marked difference regarding the ranges which each one covers. In general, the series section permits matching a much wider range of impedances than the shunt section. It is interesting to note that the shunt sections are useful precisely in the range in which the series section cannot yield a match. If an additional section of line is included, both systems can be designed to match any impedance. With regard to the method used it is seen that the introduction of the terminal functions ρ and Φ leads to great mathematical simplicity at several points. In particular, the effect of the introduction of an additional section of line would not be so obvious if these two terminal functions were not used.

The writer wishes to acknowledge his indebtedness to Professor Ronold King for suggesting this problem and guiding his work. He also wishes to express his appreciation of the helpful suggestions and criticisms of Dr. P. LeCorbeiller and of Dr. A. H. Wing, Jr. Acknowledgment is also due his friend Mr. C. M. Liang who helped to prepare most of the figures.

On the Proportion of Crystalline and Amorphous Components in Stretched Vulcanized Rubber*

A. J. WILDSCHUT

Rubber Foundation, Delft, Holland

(Received September 28, 1945)

Determinations of the tension-temperature relation of stretched vulcanized rubber can provide us with data about the proportion of crystalline and amorphous components. The amount of crystalline material appears to be 30–32 percent at 600 percent elongation and at 20°C. An expression is derived relating percentage crystalline material and temperature. The results are in close agreement to those of x-ray measurements carried out by Goppel but diverge largely from those obtained by Field. The cause of this difference is not yet quite clear. Stretched vulcanized rubber consists of a predominating amorphous phase with crystallites embedded in it. On stretching orientation occurs and a systematical addition of secondary valency forces is possible. This is the main cause of the existence of a certain tensile strength. Crystallization, though important as an indication on orientation, is more or less an incidental phenomenon. The distance function of the secondary forces may be of equal importance with the orientation.

1. INTRODUCTION

SINCE Katz¹ recognized the crystalline structure in stretched rubber, x-ray analysis has improved our knowledge about high polymeric elastic substances to a considerable extent. The influence of temperature on the deportment of the rubber crystallites was studied on raw as well as on vulcanized rubber. Various influences of mixing and vulcanization on the x-ray pattern were investigated. The fibrous character of stretched rubber was definitely established, which contributed considerably to the important new theories about high elasticity.

Long before the crystallinity of stretched and of "frozen" rubber was proved by the x-ray diagram, it was understood that physical properties such as specific weight, hardness, light-transmission showed transition points when their courses with temperature were determined. It was known for a long time that stored raw rubber could become hard and opaque, while the original suppleness could be restored by heating or even by repeated bending and stretching. Bunschoten² pointed out that the specific gravity of "frozen" raw rubber decreases considerably after heating and increases again after the rubber is kept for a

certain time at a relatively low temperature. Though the possibility of a phase-change was considered, the "freezing" of rubber was not yet then explained by presuming a partial crystallization. In later years analogous experiments, though over a much longer range of temperature, were done by Bekkedahl,³ who found several transition points in the specific volume-temperature curve.

Many other investigations could be mentioned in this connection, but those two suffice in pointing out that x-ray analysis need not be the only way in which the crystallization of rubber can be studied. After recognizing the importance of crystallization on the plastic properties of vulcanized rubber,⁴ the author made a study of the relation between elastic tension and temperature, which relation appeared also to be influenced by crystallization to a considerable extent. Though originally meant for a different purpose, this study proved to be valuable also for a *quantitative* investigation of the crystallization. It appeared possible to calculate the percentage of crystalline material in stretched vulcanized rubber from tension-temperature relations as determined for thermodynamical investigations.

* Communication number 51 of the Rubber Foundation, Delft, Holland.

¹ J. R. Katz, "Ueber die Aenderungen des Röntgenspektrums bei der Dehnung," *Kolloid Zeits.* 36, 300 (1925).

² E. Bunschoten, "Eenige waarnemingen betreffende het soortelijk gewicht van ruwe en ge vulcaniseerde rubber," *Versl. en Meded. Afd. Handel Dep. van Landb., Nijverh. en Handel* (1921) No. 6.

³ N. Bekkedahl, "Forms of rubber as indicated by temperature-volume relationship," *J. Research Bur. Stand.* 13 (September, 1934).

⁴ See A. J. Wildschut, "Pseudo-plastic phenomena with vulcanized rubber," *Physica* 10, 571 (1943). "Les phénomènes pseudo-plastiques dans le caoutchouc vulcanisé," *Rev. Gén. du Caoutchouc* 21, 9 (1944).

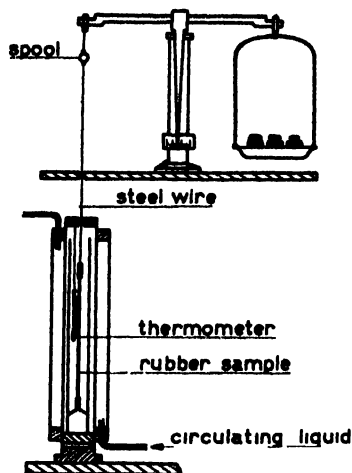


FIG. 1. Apparatus for determining the tension-temperature relation.

It is not the author's intention to deal with these latter investigations⁶ in full, as this is quite a different subject. Only the side path, leading to a new method of quantitative determination of the degree of crystallization, will be treated here.

2. THE TENSION-TEMPERATURE RELATION

The well-known thermodynamical formula

$$\sigma = (\partial U / \partial L)_T + T(\partial \sigma / \partial T)_L, \quad (1)$$

where σ is the elastic tension, U the internal energy, L the length, and T the absolute temperature, used already by Wiegand and Snyder,⁶ denotes a linear relation between the tension σ and the temperature T . The slope of the σ - T line fixes $(\partial \sigma / \partial T)_L$, while the point of intersection with the ordinate gives the value for $(\partial U / \partial L)_T$.

This formula is valid provided the process is reversible. Therefore the rubber sample under investigation must be sufficiently elastic so that plastic flow during the measurements may be neglected. With a normal vulcanized pure gum mixture from natural rubber it is comparatively easy to meet this requirement. If such a rubber is stretched, heated for some time at, for example, 70°C till the rate of relaxation is sufficiently small, cooled down to room temperature,

⁶ A survey is given in: A. J. Wildschut, *Technological and Physical Investigations of Natural and Synthetic Rubbers* (Elsevier, Amsterdam, 1945). (In press).

⁷ W. B. Wiegand and J. W. Snyder, "The rubber modulus, the Joule-effect and the dynamic stress-strain Trans. I.R.I. 10, 234 (1934).

and heated again to 70°C, as has been done by Meyer and Ferri,⁷ the original tension is reached again or is so nearly approximated that the difference may be neglected. As the process is reversible, Eq. (1) is valid and the relation found between tension and temperature must be linear. Up to a certain extent the experimental data are in accordance with the thermodynamical theory, but for a complete explanation of all experimental facts it appeared necessary to consider a possible partial crystallization of the rubber investigated. This crystallization is influenced also by the velocity with which the experiments are carried out.

3. METHOD OF DETERMINING THE TENSION-TEMPERATURE RELATION

The apparatus (see Fig. 1) comprises a part regulating the temperature of the test piece and a part by means of which the elastic tension can be measured at constant elongation. As it was believed desirable to heat the rubber in air, a double-walled glass-mantle, through which a liquid could be pumped, was constructed around the test piece, while the tension was measured by means of an ordinary balance from which one of the scales had been removed. The elongation was determined and controlled by measuring the distance of two marks on the rubber by means of two calibrated glass plates, acting as cathetometer.

The stretched rubber was even more sensitive to slight fluctuations in temperature than the precision-thermometer used. By means of two water baths, however, one for heating and cooling and the other for measuring at constant temperature, the temperature inside the mantle could be kept sufficiently constant.

Small variations in elongation of the part of the test piece between the marks could occur, as this part was not always elongated in the same way as the rest of the test piece. This could then be corrected by means of a small spool in the steel wire connecting the sample with the balance.

The cross section of the test pieces used was mostly about 0.5 mm² (e.g., 0.3 × 1.4 mm); the loads varied up to about 200 grams and could be measured with an accuracy of 0.1 gram.

⁷ K. H. Meyer and C. Ferri, "Sur l'élasticité du caoutchouc," *Helv. Chim. Acta* 18, 570 (1935).

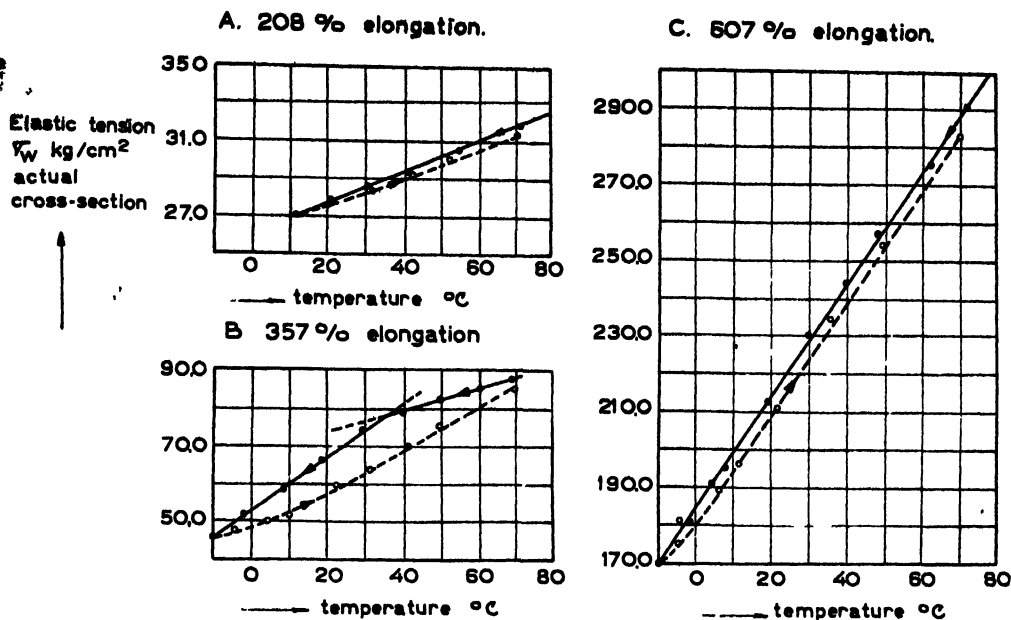


FIG. 2. Tension-temperature diagrams of vulcanized natural rubber (see reference *a* of Table I) at different elongations.

Originally the experiments were carried out over a range from room temperature to 70°C, but in the course of the investigations it proved to be necessary to extend this range up to 100°C and down to -10°C. For temperatures below room temperature, alcohol, cooled in a mixture of solid carbon dioxide and acetone, was pumped through the glass mantle.

Most experiments were carried out after heating the stretched rubber at 70°C during 4 hours. The sample was then cooled down in steps of about 10 degrees Celsius. At each temperature the rubber was kept for about 10 minutes.

If the sample was heated to temperatures above 70°C, the time of relaxation, at least in the cases of higher elongations, had to be shorter than 4 hours, as otherwise breaks might occur. As the rate of relaxation increases strongly with increasing temperature, no difficulties were presented by this.

4. INFLUENCE OF CRYSTALLIZATION

If stretched and heated vulcanized rubber is cooled down again, three cases can be distinguished, dependent on elongation and rate of cooling. Examples of these cases are given in Fig. 2.

First case, Fig. 2A. Low elongation (208 percent). There is but little crystallization, and after having been kept at 70°C during 4 hours, the rubber is certainly amorphous. On cooling, the amorphous state can be maintained and in that case the tension-temperature relation is straight, in accordance with Eq. (1). On heating, the initial point is reached again, proving that relaxation during the cycle may be neglected.

Second case, Fig. 2B. Medium elongation (357 percent). The rubber crystallizes on stretching, but after 4 hours at 70°C the sample may be considered amorphous. On cooling, the tension decreases linearly with temperature till the transition point of 36°C. Then partial crystallization sets in, causing a marked increase in the slope of the tension-temperature line, which, however, remains straight.

Third case, Fig. 2C. High elongation (607 percent). The stretched sample is partly crystallized and 70°C is too low to melt the crystallites under these circumstances. On cooling, a straight line without a transition point is obtained, as over the whole range the sample is already (partly) crystallized. The increase in degree of crystallization, which certainly takes place on cooling, increases the slope of the line but apparently does not

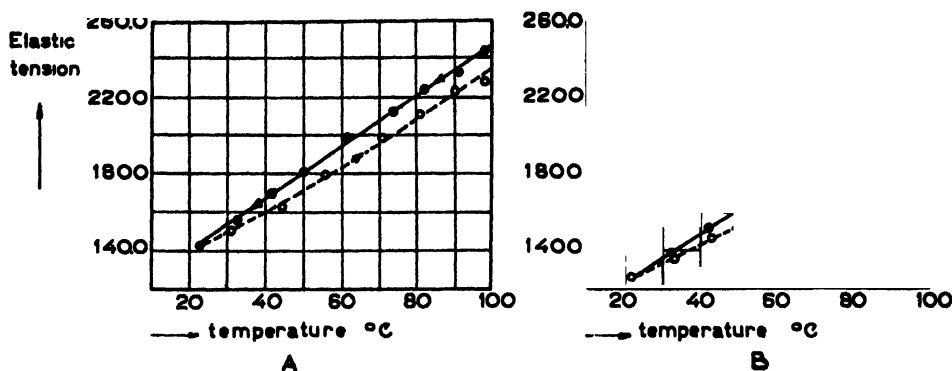


FIG. 3. The tension-temperature relation at 600 percent elongation. Influence of the time of heating. Natural rubber pure-gum vulcanizate (see reference *a* of Table I). *A*. Heated 15 min. at 100°C. *B*. Heated 60 min. at 100°C.

not influence the linear course of the tension-temperature relation.

From Fig. 2B it can be seen clearly that the tension of partly crystallized rubber is lower than it would be if no crystallization had taken place. This latter situation can easily be realized if the rubber is cooled down comparatively rapidly, as has been done by Meyer and Ferri,⁷ who had their test pieces in direct contact with the cooling medium, water. This is the reason why these investigators could not observe a transition point.

On heating again a retardation in the melting of the crystallites occurs, with the effect that partial crystallization still exists even in the range above the melting point. The observed lower tensions are a consequence of this retardation as they cannot be caused by relaxation. Ultimately the initial tension is practically reached again, which gives proof that relaxation may be neglected.

At higher elongations very little retardation appears to occur (Fig. 2C). The reason for this phenomenon is not clear as yet. Also in this case the initial tension is approximated sufficiently to justify neglect of relaxation.

5. THE "MELTING POINT" OF THE CRYSTALLITES

It was learned from measurements at medium elongations that the transition point in the σ - T curve was displaced to higher temperatures when higher elongations were used. Therefore it was logical to search for the missing transition points at higher elongations by measuring over a wider range of temperature. For this purpose the initial

point was raised to 90°C, later even to 100°C. At elongations of about 600 percent breaks often occurred, but sufficient relaxation was obtained in a short time, e.g., 15 minutes. Although the initial tension is not quite reached again (see Fig. 3A), the influence of relaxation during the cooling may still be neglected, as at lower temperatures the rate of relaxation decreases strongly.

As the curve of Fig. 3A shows no transition point, heating for 15 minutes at 100°C is not sufficient to melt a noticeable part of the crystallites. After heating during 60 minutes, however, the transition point appears at about 74°C (Fig. 3B). From further contemplations it can be derived that in this case by heating at 100°C during 60 minutes about 60 percent of the crystalline phase is melted.

So it is quite clear that at higher elongations not only a high temperature, but also a comparatively long time is needed to melt the crystallites. On cooling, crystallization sets in on a fairly definite point; there is only a small range of temperature during which the crystallization begins. After that range has been passed, the percentage crystalline material appears to increase in inverse proportion to temperature (see Section 6). The rather sharp solidification point may be caused by the existence of crystallization nuclei or small crystallites acting as nuclei.

As the melting of the crystallites appears to depend so strongly on circumstances, it is not astonishing that no distinct relation between transition points and elongation could be derived from the experimental data. Yet, when the

results of 26 measurements are plotted on squared paper, as has been done in Fig. 4, it is quite clear that a certain relation is present.

An increase in elongation from 300 to 600 percent corresponds with a displacement of the transition-point from about 40 to about 65°C.

These measurements provide us with a qualitative picture of the influence of crystallization on the tension-temperature relation. The most important conclusion is that the linear course of this relation is not disturbed by increasing partial crystallization.

6. DETERMINATION OF THE DEGREE OF CRYSTALLIZATION

In Fig. 5 the course of tension with temperature at medium elongation, as given already in Fig. 2B, is pictured again.

The rubber has been stretched, heated for 4 hours at 70°C, and then cooled down comparatively slowly. The stretch *AB* corresponds with wholly amorphous rubber (see below) and is thus in accordance with Eq. (1). Below the transition point at *B* partial crystallization occurs. The stretch *BC* therefore corresponds to partial crystallized rubber *with increasing amount of crystalline material*. The slope of this part of the curve is caused by the combination of the decrease according to form, (1), known to be linear, and the decrease, caused by the growing amount of crystalline material. The result is a linear

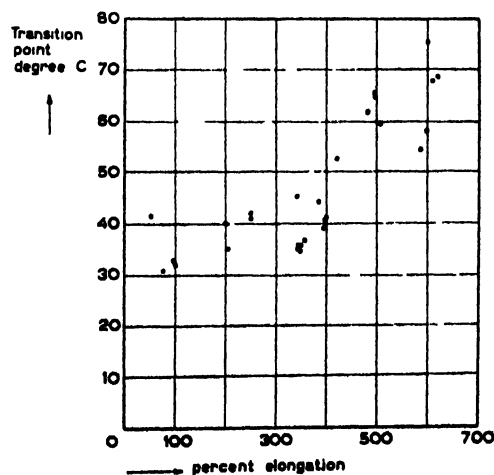


FIG. 4. Transition-points in the σ - T relation as a function of the elongation. Natural rubber pure-gum vulcanizate (see reference *a* of Table I).

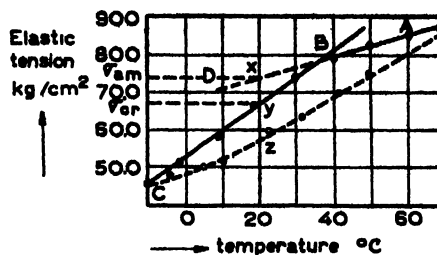


FIG. 5. Calculating the degree of crystallization from the tension-temperature relation. (Natural rubber pure-gum vulcanizate (see reference *a* of Table I) at 357 percent elongation.)

decrease of tension with temperature; therefore the decrease, caused by the crystallization-effect, must also be linear.

Now this decrease is caused by the fact that decreasing temperature causes an increasing amount of crystalline material, while more crystallization involves a lower tension. As the combination of these two functions is linear, it is very likely that both of them are linear also. If that is so, the conclusion would be reached that *the elastic tension decreases in inverse proportion to the amount of crystalline material*.

Presuming this conclusion valid, the calculation of the degree of crystallization in stretched vulcanized rubber from measurements as given in Fig. 5 is very simple.

If no crystallization had taken place, the tension-temperature relation would have followed the course *BD* and at 20°C the point *X* would be passed, corresponding with an "amorphous tension," σ_{am}^{20} , of 73.7 kg/cm². As, however, the rubber is partly crystallized, at 20°C the point *Y* is reached, corresponding with a "crystalline tension," σ_{cr}^{20} , of 66.9 kg/cm². This tension is 9.2 percent lower than the "amorphous tension," which denotes that 9.2 percent of the rubber molecules have been eliminated by crystallization.

In general, the percentage crystalline material *K* is given by:

$$K = \frac{\sigma_{am} - \sigma_{cr}}{\sigma_{am}} 100. \quad (2)$$

The tension in stretched rubber is thus caused by the amorphous phase only. Only these molecules have a possibility to return to their most

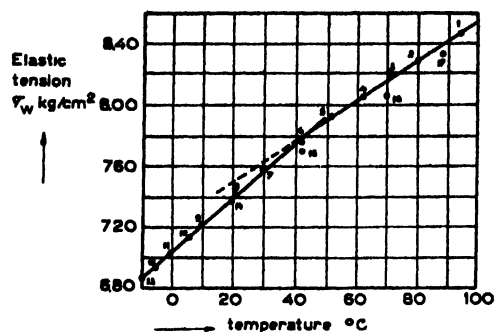


FIG. 6. Existence of crystallization in stretched vulcanized rubber (see reference *a* of Table I) at 55 percent elongation.

probable position; as soon as they have entered a crystal lattice, this possibility no longer exists.

If the test piece, after cooling to -10°C , is heated again, the tension at 20°C corresponds to that of the point Z (see Fig. 5). On passing Z, a higher degree of crystallization still exists, but this situation is not stable and is therefore left out of account. It denotes, however, the necessity to consider always the history of a given piece of rubber if its crystallization is to be measured.

Only if it is certain that at the initial point A the rubber is wholly amorphous, exact data about the degree of crystallization are obtained, for the transition point may appear also when at higher temperatures the rubber is still partly crystallized. In those cases only an increase in degree of crystallization is found and the total amount of crystalline material may be higher.

It is, however, rather simple to prove the presence of the wholly amorphous state above the transition point. If the rate of cooling is augmented, the "crystalline part" of the tension-temperature line becomes curved—for crystallization takes time—while the "amorphous part," if wholly amorphous, remains straight. In this way it has been proved that, in cases as pictured in Fig. 5, the test piece must have been wholly amorphous at 70°C .

Theoretically the "crystalline part" of the tension-temperature curve cannot be straight, as crystallization is dependent on time, but as this relation is logarithmical, it suffices to measure with the velocity described in Section 3 to get a line that cannot be distinguished from a straight one.

With x-ray measurements below 300 percent

elongation, no crystallization is observed in natural rubber pure gum mixtures. By means of the method described above, however, the existence of crystallization effects can be shown, and the amount of crystalline material can be calculated at elongations as low as 50 percent. The degree of crystallization is then naturally very small, about 1 percent, but the effect is unmistakable, as is shown in Fig. 6.

As yet it has not been possible to control all factors influencing crystallization at lower elongations. Sometimes crystallization occurs, but in other cases it may not happen at the same or even higher elongations. An example of this has already been given in Fig. 2A. The degree of crystallization found varies also in separate cases. But this is not considered essential at this stage of the investigations. Stress can be laid on the conclusion that at these low elongations a distinct, though very small, crystallization effect occurs. Whether at 50 percent elongation an exact degree of crystallization can be determined at all is not certain, as it is possible that this degree varies according to circumstances, and factors depending on the laws of probability may also play a part.

The degrees of crystallization, calculated from 25 measurements on a natural rubber pure gum mixture (different vulcanizates, but always the same mixture and cure), are given in Fig. 7 as well as in Table I.

As soon as the rubber is elongated, crystallization sets in, but up to 250 percent not more than

TABLE I. Percentage crystalline material in stretched vulcanized rubber^a at increasing elongation. 20°C .

Percent elongation	Percent crystalline material	Percent elongation	Percent crystalline material
50	1.4	347	8.6
55	1.3	350	9.5
97	0.5	357	9.2
100	0.5	385	19.6
120	0.9	390	12.6
200	0.9	392	17.7
205	0.8	397	19.3
250	2.0	400	16.0
300	4.3	420	23.8
342	9.7	480	27.4
342	8.3	492	24.5
345	10.4	495	22.8
345	7.5		

^a Composition. Crepe rubber 100, diphenylguanidine 0.5, mercapto-benzothiazole 0.8, sulphur 1.75, zinc oxide 5.0, stearic acid 1.0, and α -naphthylamine 1.0. Cure 20 minutes at 142°C in a press.

2 percent is crystallized. Above 300 percent elongation the crystallization increases rapidly, till at 500 percent a maximum of about 27 percent crystalline material is reached. By extrapolating the curve of Fig. 7 to 600 percent elongation, the maximum degree of crystallization can be estimated to be about 30–32 percent. So even in that case the amorphous phase predominates.

Presuming that at 600 percent elongation, 30 percent of the stretched rubber is crystallized, it is possible to gain complete insight in Fig. 3B. When passing at 20°C from the extrapolated upper part of the curve to the lower part, the tension decreases from 148.4 to 122.7 kg/cm², that is, with 17.4 percent. If the upper part had represented the wholly amorphous state, the decrease would have been 30 percent. So $17.4 \cdot 100 / 30.0 = 58$ percent of the crystalline phase was melted by heating at 100°C during an hour.

Of course in such a calculation this exactness has no sense, but the conclusion may be drawn that rather more than half of the crystalline phase will be melted in an hour and that it will take at least another hour to reach the wholly amorphous state. Unfortunately most samples cannot stand this, so 600 percent is about the natural limit for this sort of investigation. Perhaps an improvement can be made by heating the unstretched sample at 100°C and then elongating it; these experiments will be done in due course.

7. INFLUENCE OF TEMPERATURE ON DEGREE OF CRYSTALLIZATION

As the relation between tension and temperature is linear, a tension coefficient at constant elongation, $(\partial\sigma/\partial T)_L$, can be calculated from these measurements. As long as the rubber is amorphous, the value of this tension coefficient depends on the elongation only. Below the transition point the tension coefficient increases, for the increasing degree of crystallization causes an additional decrease in tension. The "crystalline tension coefficient" is therefore dependent on time in the same way as the crystallization itself.⁴

The degree of crystallization can be calculated from the tension coefficients and the transition temperature as follows:

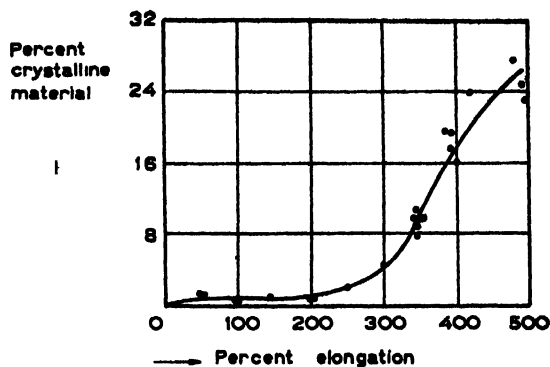


FIG. 7. Percentage crystalline material in stretched vulcanized natural rubber (see reference *a* of Table I) at 20°C.

From Fig. 5 it can be seen that the tension of the transition point, σ_{T_s} , is given by:

$$\sigma_{T_s} = \sigma_{T_{am}} + (\partial\sigma_{am}/\partial T)_L(T_s - T), \quad (3)$$

as well as by:

$$\sigma_{T_s} = \sigma_{T_{cr}} + (\partial\sigma_{cr}/\partial T)_L(T_s - T), \quad (4)$$

where $\sigma_{T_{am}}$ and $\sigma_{T_{cr}}$ are, respectively, "amorphous" and "crystalline" tensions at a temperature T and $(\partial\sigma_{am}/\partial T)_L$ and $(\partial\sigma_{cr}/\partial T)_L$ are the tension-coefficients for the wholly amorphous and partially crystallized rubber, respectively.

If the values for $\sigma_{T_{am}}$ and $\sigma_{T_{cr}}$, derived from Eqs. (3) and (4), are substituted in Eq. (2), we get:

$$K_T = \frac{[(\partial\sigma_{cr}/\partial T)_L - (\partial\sigma_{am}/\partial T)_L](T_s - T)}{\sigma_{T_s} - (\partial\sigma_{am}/\partial T)_L(T_s - T)} 100, \quad (5)$$

from which the percentage crystalline material K_T can be calculated at any temperature T . Equation (5) can be written $xy + ax + by - c = 0$, which is an equation of a hyperbola. As can be seen from Fig. 8, in practice the relation between K_T and T is approximately linear, as long as short ranges are considered.

The amounts of crystalline material have been calculated for the elongations 357 (the case represented in Fig. 5) and 492 percent; the results are given in Fig. 8.

The straightness of the tension-temperature curve has been proved down to -10°C. Below this temperature, calculation of K_T from Eq. (5) will have little or no physical meaning. In the first place the rate of crystallization diminishes

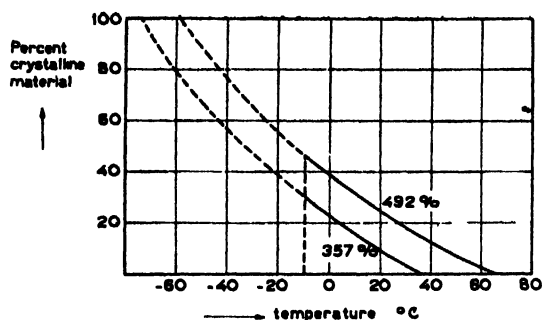


FIG. 8. Percentage crystalline material as a function of temperature. Natural rubber pure-gum vulcanizate (see reference *a* of Table I).

at decreasing temperature so strongly that for this reason alone it is doubtful whether the percentages represented by the dotted part of the curve will ever be reached at all. Secondly Eq. (5) is only valid as long as the "crystalline part" of the $\sigma-T$ curve remains straight as well as continuous. Any change in modification will cause a break in the $\sigma-T$ curve. This makes an extrapolation below the temperatures of the measurements without physical significance.

The degree of crystallization appears to be strongly dependent on temperature. While at room temperature (20°C) the maximum percentage crystalline material may be estimated to be about 30 percent, at zero degrees this maximum may be 45 percent and at -10°C even as high as 50 percent. The highest temperature at which crystallization in stretched rubber exists appears to be about 80°C, a conclusion that can be derived also from Fig. 4. Above 60°C the amount of crystalline material remains below 10 percent, provided the situation is stable. Of course, if stretched rubber is heated rather rapidly, higher degrees may temporarily occur, but then the system is not in equilibrium.

8. COMPARISON WITH X-RAY MEASUREMENTS

Quantitative measurements of the crystallization in stretched vulcanized rubber by means of the x-ray method have been done by Field⁸ and by Goppel.⁹ Both investigators compared

the intensities of the diffuse rings in the diagrams of amorphous and partly crystallized rubber, while also measurements of the intensities of the diffraction spots were done.

It lies beyond the scope of this paper to discuss these investigations in full. As it happens, however, the results of Field and of Goppel show a large difference: while the former finds at 500 percent elongation and at room temperature a degree of crystallization of 65–70 percent, the latter measures 25–30 percent, which is in accordance with the results of the thermodynamical method described above.

The work of Goppel was carried out independently of that of the author and most of it at an earlier date. His results as given in Fig. 9 are obtained on the same mixture on which the thermodynamical measurements have been done, but also various other pure-gum mixtures have been investigated,¹⁰ and it appeared that the differences between them were small.

Results of both x-ray investigations, as well as the results obtained with the thermodynamical method, have been collected in Fig. 9.

There is a close agreement between the results of Goppel and those of the author, while the data given by Field diverge very considerably. It is most unlikely that this is caused by the comparatively small differences in mixture and cure, but it is not yet possible to state the exact cause of this phenomenon.

Various arguments can be advanced for the existence of not more than a moderate amount of crystalline material in stretched vulcanized rubber. It is unlikely that an entangled knot, with primary valency bonds between the molecules, can be stretched in such a way that a great amount of regularity is gained. As soon as certain parts of the chains have entered into crystal lattices, other parts are hindered in their movements and less able to do the same. The primary valency bonds caused by the vulcanization arise while the rubber is unstretched; by stretching some of these bonds may break,

⁸ J. E. Field, "An x-ray study of the proportion of crystalline and amorphous components in stretched rubber," *J. App. Phys.* **12**, 23 (1941).

⁹ J. M. Goppel, "Quantitative Röntgenografische Untersuchungen an Rubber," Thesis, Delft, 1945. (In German.)

¹⁰ The results of Goppel and of the author have been obtained on the vulcanizate, defined in note *a* of Table I. The data of Field have been derived from his Figs. 10 and 11. Compound B: smoked sheet 100, zinc oxide 2, sulphur 2, P.P.D. 2, cure 10 minutes at 260°F (127°C). Compound C: smoked sheet 100, zinc oxide 1, sulphur 2, zincate 1, cure 20 minutes at 260°F.

but others will not. It is in contravention of probability that the irregular network can be transferred in a system of crystal lattices; so crystallites can be formed only to a limited amount.

The thermodynamical measurements can only be compared with the two x-ray investigations mentioned. One is in close agreement with them, the other diverges largely. A third, if possible independent, measurement of the degree of crystallization will be required to make certain where the truth lies.

9. GENERAL PICTURE OF STRETCHED VULCANIZED RUBBER

The author is inclined to believe that the lower figures for the amount of crystalline material in rubber are right. This leads to the assumption that stretched vulcanized rubber is composed of an amorphous phase with up to a third part crystallized material. The existence of crystallization is important as an indication on regularity; *but it is doubtful whether crystallization, in itself, is of paramount importance for the tensile strength of rubber.*

Immediately after the elongation begins, orientation of the rubber molecules occurs. This orientation will be one of the main causes of the formation of a high mechanical firmness

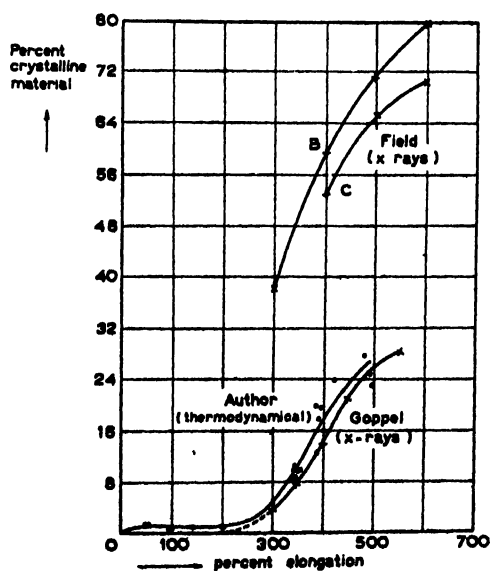


FIG. 9. Crystallization in stretched vulcanized rubber (see reference 10) according to three investigations.

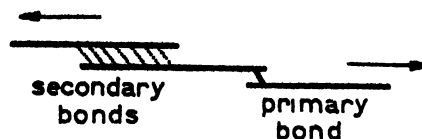


FIG. 10. On the mechanism of break.

during the elongation. Crystallization, though directly important for certain other properties, is only an incidental phenomenon and it is not even certain whether the existence of crystallites adds to the tensile strength or not. At any rate a high tensile strength is very easily possible without any crystallites playing a role, also in pure-gum vulcanizates. At temperatures up to 100–140°C vulcanized rubber keeps its tensile strength completely,^{11 12} while certainly under these circumstances no crystallization is present.

Of equal importance with regard to the tensile strength may be the way the secondary forces between the molecules decrease at increasing mutual distance. If this function is favorable, a high tensile strength may occur even at medium orientation.

It is not likely that the primary valency bonds in vulcanized rubber are decisive for the tensile strength. They are, of course, important with regard to plastic flow, but it is unlikely that, on stretching, they are responsible for equal parts of the tension. So in the region shortly before break these primary bonds may be broken one after another, if the combined secondary energies between two chain-parts, drawn parallel, are higher than the energy of the primary bond concerned. (Compare Fig. 10.)

The tensile strength will be caused mainly by secondary forces, added systematically along the chains. This addition is possible for secondary forces only.¹³ The crystallites may exert additional forces on the surrounding amorphous phase and thus add to modulus and tensile strength, in the same way as carbon black in

¹¹ A. van Rossem and H. van der Meijden, "Studies on the physical properties of rubber. I. Influence of high temperature on the stress-strain curve of vulcanized rubber," *J. Soc. Chem. Ind.* **45**, 67T (1926).

¹² A. A. Somerville and W. F. Russell, "Effect of temperature on tensile properties of vulcanized rubber," *Ind. Eng. Chem.* **25**, 1096 (1933).

¹³ See also G. Briegleb, *Zwischenmolekulare Kräfte und Molekülstruktur. Sammlung chemischer und chemisch-technischer Vorträge*, New edition, No. 37 (1937).

Perbunan does, but it is equally possible that this effect—if it exists at all—is overcompensated by the partial elimination of rubber molecules by the crystallization (which causes, as has been shown above, a decrease in elastic tension). Of course the crystallites in themselves are firmly bound together, but this is of no importance to the tensile strength, as they are embedded in amorphous material. This may be the most important conclusion from our measurements about the degree of crystallization: when determining the tensile strength *the forces are applied on an amorphous phase* and not on a bundle of crystals.

This vision on the situation in stretched vulcanized rubber is strengthened by various phenomena, observed on synthetic rubbers,⁵ with and without carbon black. Synthetic rubbers with very weak pure-gum mixtures show a normal high tensile strength when carbon black

is added: secondary forces must play a role. Lightly swollen rubber¹⁴ has lost most of its mechanical firmness; after evaporation of the swelling liquid the original tensile strength is completely restored. This phenomenon can only be explained by presuming secondary forces to be responsible for the tensile strength.

More could be said about this, but that would carry us much too far. It does not lie within the scope of this article to discuss all physico-chemical aspects of the rubber molecule, but only to draw attention to the fact that a relatively low degree of crystallization is likely to exist and leads to an essentially different picture of stretched rubber from that of a high degree. The picture, corresponding with a low degree of crystallization, agrees better with phenomena observed on synthetic rubbers.

¹⁴ A swelling of 5–10 volume percent will be enough.

Letters to the Editor

On Plane Rotations in n Dimensions

PETER T. LANDSBERG
London, England
October 11, 1945

THIS note is intended as a contribution to the discussion of the "Representation of Rigid Rotations" which has recently been taking place in this journal.¹ The main formula (Eq. (1) below) derived by Professor Benedikt is an operational equation and will be shown in this note to be a three-dimensional statement of equations known to hold in n dimensions. Use will be made of the concepts and nomenclature of modern algebra.²

Consider a linear space γ_n of finite dimensionality n (≥ 3) and defined over a non-modula field $\mathcal{F}$. Elements of γ_n , i.e., vectors, will be denoted by small Greek letters (θ is reserved to denote a variable angle), operators by capital Latin letters, and elements of $\mathcal{F}$ by small Latin letters. In γ_3 the ordinary vector product ($\times$) of two vectors is assumed to be defined.

Let any vector α_0 be rotated through a variable angle θ , yielding a vector α_θ , subject to the requirement that α_0 and all α_θ 's are co-planar. Let the point about which α_0 is rotated be specified by a radius vector ρ relative to a fixed origin O . The end points of α_0 and α_θ may then be specified relative to O by radius vectors

$$\rho_0 = \rho + \alpha_0, \quad \rho_\theta = \rho + \alpha_\theta,$$

respectively. Further, let ρ be perpendicular to the plane defined by the α_θ 's. Then Professor Benedikt has shown that for $n=3$

$$\rho_\theta = e^{\theta \gamma} \rho_0, \quad (1)$$

γ being the unit vector parallel to ρ and in a positive direction relative to the rotation.

Consider now any linear space γ_n . Let an orthogonal normalized basis $\{\epsilon_k\}$ be given in γ_n such that α_0 may be expressed in terms of ϵ_{10} and ϵ_{20} , ρ in terms of $\{\epsilon_k\}$ ($k=3, 4, \dots, n$). The n dimensional analog of Eq. (1) may then be obtained by assuming α_0 to be fixed but re-expressing it in terms of a new basis $\{\epsilon_{j\theta}\}$ obtained by rotating ϵ_{10} , ϵ_{20} through $(-\theta)$. The expression for α_0 in terms of $\{\epsilon_{j\theta}\}$ will obviously be the same as the expression for α_θ in terms of $\{\epsilon_{j\theta}\}$. If θ is continuously variable and $\mathcal{F}$ the field of all-complex numbers, a group of unitary operators $U(\theta)$ (these reduce to real orthogonal operators in the case of the real field) may be supposed to induce the transformation $\{\epsilon_{j\theta}\} \rightarrow \{\epsilon_{j0}\}$. If Y be an anti-symmetric operator, $U(\theta)$ may be represented in the form $e^{\theta Y}$, the latter expression being defined by the exponential series.³

³ It is not necessary to invoke this property of unitary matrices if some equation of the type $\epsilon_{j\theta} = U(\theta)\epsilon_{j0}$ is assumed possible. For it then follows from the nature of the problem that the $U(\theta)$'s have the property

$$U(\theta_1) \cdot U(\theta_2) = U(\theta_1 + \theta_2), \\ U(0) = U(2\pi) = \text{identity operator},$$

whence $U(\theta)$ has an exponential representation. This is in fact merely a discussion of the representations of a one-parameter continuous Abelian group,

¹ J. Benedikt, J. App. Phys. 15, 613 and 892 (1944).
² For instance, A. A. Albert, Introduction to Algebraic Theories (Chicago, 1941).

Hence

$$e_{j0} = e_{j0}^Y \quad (j=1, 2, \dots, n); \quad e_{k0} = e_{k0}^Y \quad (k=3, 4, \dots, n);$$

whence

$$\alpha_0 = e^{\beta Y} \alpha_0, \\ \rho = e^{\beta Y} \rho_0.$$

Therefore,

$$\rho_0 = e^{\beta Y} \rho_0. \quad (2)$$

Defining:

$$\lim_{\delta\theta \rightarrow 0} \frac{\rho_0 + \delta\theta - \rho_0}{\delta\theta} = \frac{d\rho_0}{d\theta},$$

we also have

$$\frac{d\rho_0}{d\theta} = Y\rho_0. \quad (3)$$

Equations (2) and (3) hold in n dimensions. In γ , Y is easily identified from (3) as the operator ηx so that (2) reduces to (1) in this case.

Letter to the Editor

GERALD PICKETT

Kansas State College, Manhattan, Kansas

October 5, 1945

THE purpose of this letter is to call attention to the trend of many contemporary writers to regard surface tension as something not quite real, and to present a few arguments against this view. For example, in discussing properties of liquid surfaces Glasstone¹ states "The tendency for a liquid to contract may be regarded as a consequence of its possession of free energy; . . . As a result of the tendency to contract a surface behaves as if it were in a state of tension." Adam² is more positive that surface tension does not really exist, that it only serves as a mathematical device. He says, "The essential mechanism, . . . of the production of the free surface energy, is the perpendicular inward attraction, exerted on the surface molecules by the underlying ones, there is no need to speculate how this can be transformed into a surface tension parallel to the surface, for the surface tension does not exist as a physical reality, and is only the mathematical equivalent of free surface energy."

The writer believes that surface tension is real, as real as any stress that may exist in a solid, liquid, or gas, and is concerned that progress in research may be hampered by the contrary view. A few arguments in favor of the writer's view follow.

For simplicity, consider only the surface tension of liquids at liquid-gas interfaces. For surface tension to exist points within the surface layer must have unequal principal stresses. Of course it is well known that at any point within a homogeneous liquid or gas at rest, the three principal

stresses are equal. But a vapor-liquid interface is not a homogeneous phase, and, as a consequence, the principal stress normal to the interface may differ from the other two stresses.

Surface tension γ may be defined as

$$\gamma = \int (\sigma_t - \sigma_z) dz, \quad (1)$$

where σ_t and σ_z are the tangential and normal stresses, respectively, and the integration is taken over the thickness of the transition layer.

From a consideration of the equilibrium of a portion of the transition layer, the following well-known formula is readily obtained,

$$\gamma[(1/r_1) + (1/r_2)] = p, \quad (2)$$

where r_1 and r_2 are the principal radii of curvature of the interface, and p is the difference in the normal stress on the two sides of the interface.

The validity of Eq. (2) is not questioned by those who deny the reality of surface tension; they merely derive it from consideration of energy (see page 13 of reference 2) rather than from the equilibrium of a free body.

The writer will agree that for many if not most problems of interfaces the concept of surface energy will lead more directly to a solution than will the concept of surface tension just as many problems in mechanics are more simply solved by considering energy relations than by analysis of free bodies. For example, the lateral vibration of a bar could be studied by either energy methods or by methods based on free-body considerations. It might be argued (falsely) that there was no real longitudinal stress in the vibrating bar, since the motion and consequently the inertia forces are in a lateral direction. It is only from free-body analysis that one finds that longitudinal stress must accompany the lateral vibration of a bar. And from a free-body analysis of the transition layer between vapor and liquid, however thin that layer may be, one must accept the reality of surface tension.

Not only is surface tension in the transition layer between the homogeneous liquid and vapor phases real, but by making simplifying assumptions in regard to inter-particle forces; i.e., treat all forces as though they were Van der Waals forces between non-polar molecules, the thickness of the transition layer and the distribution of stress within it can be theoretically determined. An approach in this direction was made by Hulshof³ who showed that to a first approximation

$$\sigma_t - \sigma_z = \frac{C(\partial\rho)^2}{2(\partial z)^2} - \frac{C\rho\partial^2\rho}{2\partial z^2}, \quad (3)$$

where ρ is density and

$$C = \int_0^\infty s^2\psi(s)ds,$$

where $\psi(s)$ is the attraction between a semi-infinite body of the material of unit density and a body of the material of unit mass located s distance from the semi-infinite body.

¹ Glasstone, *Textbook of Physical Chemistry* (D. Van Nostrand Company, New York, 1940), p. 473.

² Adam, *Physics and Chemistry of Surfaces* (Oxford Press, New York, 1930), p. 6. (See also second edition, 1938 and third edition, 1941).

³ Hulshof, "Ueber die Oberflächenspannung," *Ann. d. Physik* [4] 4, 165 (1901).

On a Novel Form of Refrigerator

JENNY E. ROSENTHAL
Long Branch, New Jersey
October 30, 1945

THE September, 1945 issue of the *Journal of Applied Physics* contains a letter by Moody¹ criticizing "A Novel Form of Refrigerator" by Roebuck.² In addition to showing the flaws in analysis and design of the apparatus, the letter brings up several other points, which it is proposed to try to clarify in the following remarks.

As shown by Moody, Roebuck's machine is of awkward construction and cannot possibly make use of high speeds of rotation. In a subsequent publication it will be shown how a device can be designed so as to overcome these difficulties. For the present it will be assumed without proof that proper design makes it possible to raise the peripheral speed of rotation of the centrifuge at least to the value of 1300 ft./sec. given by Stodola.³

According to Kapitza⁴ the efficiency of a turbine-like cooling device rests on the fact that the gas enters the cooling chamber at high density. If the compressor is external to the turbine, then it is impractical to attain the high density by high pressure, hence Kapitza has to introduce the gas into the turbine at a low temperature. In a centrifuge, where a compressor and cooling expansion chamber can constitute a single rotor unit, it is practical to introduce the gas into the rotor at moderate pressures and hence into the expansion chamber at high pressure with

accompanying high density. However, this high pressure can only be attained by rotating the centrifuge near the limit set by Stodola. Hence the need for a good mechanical design mentioned above.

While unsatisfactory mechanically, Roebuck's device is nevertheless based on a sound underlying principle, which is different from that utilized in the Nernst gas turbine or in the device patented by Kapitza. Roebuck's reference to an article by Beams⁵ reveals a fundamental fallacy in his conception of the operation of his cycle. Beams refers to an altogether different cooling effect, namely the cooling action of air jets external to the rotor.

As pointed out by Kapitza,⁴ the gas expansion turbine proposed by Nernst is ineffective because no provision is made to introduce the gas into the device at high density. When gas at high density is used in a properly designed device constructed on the same principle as Roebuck's, the device functions as both a gas expansion and a hydraulic turbine, which makes use of centrifugal force. As mentioned before, Kapitza's device requires the introduction of a gas at a low temperature, a requirement which limits the range of usefulness of the device; although very efficient as a single stage unit, Kapitza's device is not constructed to give a large temperature change. Also, as pointed out by Moody, a single stage centrifugal unit cannot achieve a large temperature drop even if it runs at peripheral speeds near the limit set by Stodola, because the gas must expand under a pressure differential added to that attained through centrifugal force. This additional pressure differential is achieved by introducing the gas into the rotor at a different pressure from that existing at the exhaust.

¹ A. M. G. Moody, *J. App. Phys.* 16, 551 (1945).

² J. R. Roebuck, *J. App. Phys.* 16, 285 (1945).

³ A. Stodola, *Steam and Gas Turbines* (McGraw-Hill Book Company, Inc., New York, 1927), p. 405.

⁴ P. Kapitza, U. S. patent 2 280 585.

⁵ J. Beams, *Science in Progress* (Yale University Press, New Haven, 1940), Second Series, p. 246.

Erratum: Secondary Emission of Pyrex Glass

[*J. App. Phys.* 16, 453 (1945)]

C. W. MUELLER

George Eastman Laboratories of Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts

TABLE I erroneously gave a very early composition of Pyrex brand glass. The composition tested was:

TABLE I.

SiO ₂	80.5			Na ₂ O	3.8
B ₂ O ₃	12.9			K ₂ O	0.4
		Al ₂ O ₃	2.2		

Elastic Limit for Dynamic Loading

V. BARGMANN
School of Mathematics, The Institute for Advanced Study, Princeton,
New Jersey

AND

H. F. LUDLOFF
Department of Aeronautical Engineering, New York University, New
York, New York
November 25, 1945

IN dealing with dynamic processes, one is often confronted with the problem whether the elastic limit of materials can be characterized by a critical stress value, such as exists for static loading. In the case of transversal impact of a mass, striking upon the center of a simply supported beam, there are definite theoretical indications that the elastic limit is rather determined by a critical impulse, $\int_0^t F(t_1) dt_1$, where $F(t)$ is the contact force between striking mass and beam. $F(t)$ itself as well as local stresses may for short instants rise far beyond values that are permissible in statics.

Recent experiments yield precise time records of the center deflection $y_c(t)$ of a beam under impact loading. If the deflection of the beam is considered as forced vibration, it is evident that there must be a unique relation between the deflection of the beam and the contact force producing it. While the derivation of $y_c(t)$ from a known $F(t)$ is a routine calculation, the inverse problem involves Laplace transforms and "convolution" of functions. To provide a rigorous method for the experimental determination of $F(t)$ -curves, we have solved the following problem:

Assume as experimentally known the center deflection of a simply supported beam with mass M and fundamental frequency ω :

$$y(t) = \frac{1}{M\omega} \sum_{n=1,3}^{\infty} \int_0^t F(t_1) \frac{\sin n^2 \omega(t-t_1)}{n^2} dt_1.$$

What is the contact force between striking mass and beam $F(t)$?

Solution:

(1) Put

$$\sum_{n \text{ odd}} \frac{\sin n^2 \omega t}{n^2} = K(t),$$

then

$$y(t) = \frac{1}{M\omega} \int_0^t F(t_1) \cdot K(t-t_1) dt_1.$$

Let

$$\Phi(s) = \int_0^{\infty} e^{-st} \cdot \frac{F(t)}{M\omega} dt, \quad L(s) = \int_0^{\infty} e^{-st} \cdot K(t) dt,$$

$$\eta(s) = \int_0^{\infty} e^{-st} \cdot y(t) dt;$$

then:

$$\eta(s) = L(s) \cdot \Phi(s), \quad \text{i.e., } \Phi(s) = \eta(s)/L(s).$$

Hence

$$F(t) = \frac{M\omega}{2\pi i} \int_{c-i\infty}^{c+i\infty} ds \cdot e^{st} \cdot \Phi(s) = \frac{M\omega}{2\pi i} \int_{c-i\infty}^{c+i\infty} ds \cdot e^{st} \cdot \frac{\eta(s)}{L(s)}$$

can be evaluated by complex integration, if the zeros of $L(s)$ are known.

(2) One can show that

$$L(s) = \frac{\pi^4}{16\omega^2} \left[\frac{\sinh \zeta - \sin \zeta}{\cosh \zeta + \cos \zeta} \right],$$

where $\zeta = \pi(s/2\omega)^{1/2}$.

The zeros of $L(s)$ can be computed from simple transcendental equations with any required accuracy; the asymptotic values are $s_k \sim \pm i\omega(2k + \frac{1}{2})^2$, for integer $k \geq 1$. The asymptotic expression for s_k yields sufficiently accurate values, even for s_1 .

(3) Since at the start of the impact $F(t)=0$, we may assume $y(0)=y'(0)=y''(0)$; but $y''' \neq 0$. Then

$$\eta(s) = \int_0^{\infty} e^{-st} \cdot y(t) dt = \frac{1}{s^3} \int_0^{\infty} e^{-st} \cdot y'''(t) dt,$$

and

$$F(t) = \frac{M\omega}{2\pi i} \int_{c-i\infty}^{c+i\infty} ds \frac{e^{st}}{s^3 \cdot L(s)} \left[\int_0^{\infty} e^{-st_1} \cdot y'''(t_1) dt_1 \right].$$

After exchanging integrations and upon evaluating the residues according to standard methods, one obtains:

$$F(t) = M\omega \left\{ \frac{96\omega}{\pi^4} y(t) + \frac{32}{\pi^2} \sum_{n_k^*} (1 + \epsilon_{n_k^*}) \times \int_0^t \sin [\omega n_k^{*2} (t-t_1)] y''(t_1) dt_1 \right\},$$

which solves our problem. For all practical purposes: $\epsilon_{n_k^*} = 0$; $n_k^* = (2k + \frac{1}{2})$, $k \geq 1$. The series converges, as can be seen by partial integration, at least as $\sum \frac{1}{n_k^{*2}}$.

(4) We are particularly interested in the transmitted impulse:

$$\int_0^t F(t_1) dt_1 \sim M\omega \left\{ \frac{96\omega}{\pi^4} \int_0^t y(t_1) dt_1 + \frac{32}{\pi^2} \sum_{n_k^*} \int_0^t \sin [\omega n_k^{*2} t_1] y'(t_1) dt_1 \right\}.$$

Also here the series converges, at least, as $\sum \frac{1}{n_k^{*2}}$; more over $y'(t)$ and $\int_0^t y(t_1) dt_1$, respectively, can be obtained more accurately from records than $y''(t)$ and $y(t)$, respectively.

Example: For a contact force $F(t)=t$, we have the deflection:

$$y(t) = \frac{1}{M\omega} \sum_{n \text{ odd}} \frac{1}{n^2} \left\{ t - \frac{1}{n^2 \omega} \sin n^2 \omega t \right\}.$$

Inserting this in our inversion formula, we get:

$$F(t) = \frac{96}{\pi^4} \cdot t \cdot \sum_{n \text{ odd}} \frac{1}{n^4},$$

which equals t .

Here and There

New Appointments

Dr. Ivan A. Getting has been appointed associate professor in the department of electrical engineering of the Massachusetts Institute of Technology. During the war he was head of a division of the Radiation Laboratory at the Institute and directed research and development in the general field of the applications of radar to gunnery and associated problems.

Dr. Ira S. Bowen, professor of physics at the California Institute of Technology, has been appointed director of the Mount Wilson Observatory. He succeeds Dr. Walter S. Adams, who retired January 1, 1946, after serving as director since 1923.

Chester A. Snell, until recently associated with the Chemical Development Division, Aluminum Company of America, has joined the staff of Foster D. Snell, Inc., consulting chemists and engineers, Brooklyn, New York.

Paul D. Zottu of Newton Centre, Massachusetts, formerly chief engineer of Thermex Division, the Girdler Corporation, Louisville, Kentucky, has entered the field of consulting industrial electronic engineering.

The British Columbia Industrial and Scientific Research Council has announced the appointment of three physicists to its staff. Dr. S. E. Maddigan, formerly head of the Division of Physics and Electro-chemistry at the Chase Brass and Copper Company, has been appointed Director of the Council. Dr. A. C. Young, who was with the Physics Section of the National Research Council of Canada during the war, has been appointed head of the Council's Physics Division, and Dr. H. F. Batho, recently assistant professor of physics at the University of Manitoba, has been appointed Associate Research Physicist.

A. C. Menzies has been appointed Controller of Research and Development for the firm of Adam Hilger, Ltd., optical instrument makers of London.

Guggenheim Awards

Among the 36 recipients of fellowship awards announced in October by the John Simon Guggenheim Memorial Foundation were the following: Dr. Harold F. Blum, Principal Biophysicist, Naval Medical Research Institute, Bethesda, Maryland; Dr. Roy Overstreet, Assistant Soil Chemist, Division of Soils, University of California; Walter H. Pitts, staff member, Radiation Laboratory, Massachusetts Institute of Technology; Dr. Charles Kittel, Supervisor, Submarine Operations Research Group, U. S. Navy; Dr. Lindsey Helmholz, Assistant Professor of Chemistry, Dartmouth College.

National Research Council Fellowships

Science reports the announcement of the National Research Council that fellowships in mathematics, astronomy,

physics, chemistry, geology, paleontology, physical geography, zoology, botany, agriculture, forestry, anthropology, and psychology will be available for the year 1946-1947. These fellowships are awarded as a rule to persons under 35 years of age who are citizens of the United States or Canada, and who have met all the requirements for the doctor's degree. Applications must be filed on or before December 31 on forms obtainable from the Secretary of the Fellowship Board in the Natural Sciences, National Research Council, 2101 Constitution Avenue, Washington 25, D. C. A handbook describing the fellowships (stipends, conditions, and tenure) will be furnished upon request.

Dr. Jeffries Honored

Dr. Zay Jeffries of the General Electric Company has been awarded the 1946 John Fritz Medal for "leadership in the solution of problems affecting the production, conservation, substitution, and scientific appraisal of metals and alloys."

Annual Meeting of A.S.H.V.E.

The Council of the American Society of Heating and Ventilating Engineers has announced that the fifty-second annual meeting of the Society will be held January 28-30 at the Hotel Commodore in New York City.

Activities at Polytechnic Institute of Brooklyn

A symposium on "The Application of Polymer Chemistry of Textile Fibers" was held at the Polytechnic Institute of Brooklyn on November 24 under the chairmanship of Dr. Milton Harris. A series of eight seminars on "Recent Progress in Biochemistry," beginning in November, 1945, is being held at the Institute during the current academic year, and a series of fifteen seminars on alternate Saturdays, also beginning in November, is dealing with the newest developments in the scientific investigation of the chemistry of high polymers and the recent practical applications of polymers in the plastics industry.

Material Handling Society Organized

In Pittsburgh on October 8 a group of men interested in material handling voted to form a Society. Objectives for the Society include: furthering the application of good material handling practices; encouragement of an interchange of ideas among the members; promotion of education and training in the science and practice of practical coordinated material handling; arranging for the collection and dissemination of information relating to all phases of material handling. Temporary Chairman is T. O. English, Aluminum Company of America. Temporary Secretary is Richard Rimbach of Materials Publishing Company. Further information may be obtained from Mr. Rimbach at 1117 Wolfendale Street, Pittsburgh 12, Pennsylvania.

Organization of Applied Spectroscopy Society

The Society for Applied Spectroscopy has been formed by a group of practicing spectroscopists working in the New York metropolitan area. Meetings are held the first Tuesday evening of each month and consist of either a

lecture by some prominent speaker or a symposium on a general topic. At its first two meetings the Society heard lectures by Dr. O. S. Duffendack of the North American Philips Company and Dr. G. H. Dieke of Johns Hopkins University. Among the topics to be covered in symposia at coming meetings are: types of excitation, types of equipment, measurement of radiant energy, absorption spectra, and spectroscopic nomenclature. Membership in the Society is open to anyone having an interest in applied spectroscopy. Those interested can obtain more information from the Secretary, Charles L. Guettel, Driver-Harris Company, Harrison, New Jersey.

Meeting of Optical Society

The winter meeting of the Optical Society of America will be held at the Hotel Statler in Cleveland, Ohio, on March 7-9, 1946.

Industrial Research Institute at Chattanooga

The Industrial Research Institute has been established at the University of Chattanooga "to provide the following contract services on a non-profit basis to Chattanooga, the South, and the Nation: research for industry; fundamental studies in pure and applied science; the training of scientists and skilled workers; the maintenance of an up-to-date science library with reference service." Director of the new Institute is Dr. Raymond B. Seymour.

Columbia University Honored for Atomic Research

Science reports the recent presentation by the War Department to Columbia University of a scroll in recognition of its contribution to research on the atomic bomb. It

was accepted for the university by Dr. George Pegram, chairman of the University Committee on War Research and dean of the graduate faculties. Dr. Nicholas Murray Butler, president emeritus of the university, gave the closing address at the ceremony. He singled out for special commendation Dean Pegram; Dr. Harold Clayton Urey, professor of chemistry; Dr. Enrico Fermi, professor of physics; and Dr. John R. Dunning, professor of physics.

Technical Translations

The R. T. P.'s translations listed below are available from the Durand Reprinting Committee in care of California Institute of Technology, Pasadena 4, California.

These were selected for reprinting from the British Ministry of Aircraft Production R. T. P. List No. 123.

- | | |
|---------------|---|
| 2503 | <i>The Lorin Nozzle for Aircraft Propulsion</i> (Focke-Wulf Flugzeugbau, O.m.b.H. Bremen). (40 pages) |
| 2506 | <i>Planing on a Water Surface</i> (T.W.F. No. 4-5, 1940). (28 pages) |
| Sedow, L. I. | <i>Pieno-Electric Microphone of Small Dimensions</i> (Z.V.D.I. Vol. 87, No. 25/26, June 26, 1943, p. 406). (1 page) |
| 2507 | <i>Development of the Weibel Welding Process for Aircraft Construction</i> (Junkers. Nachrichten. Vol. 14, No. 7/8 July-Aug., 1943, pp. 53-59). (8 pages) |
| F. L. | <i>Laval Nozzles for Uniform Super-Sonic Flow</i> (Z.V.D.I. Vol. 84, No. 45, 9.11.40, pp. 857-862). (9 pages) |
| 2508 | <i>German De-Icing Technique</i> (Reports on development, devices, etc.). (8 pages) |
| Heilbing, F. | <i>The Role of Inhibitors in the Electrolytic Precipitation of Metals</i> (Part II-Part I, see R.T.P. Trans. 2213). (Kolloid-Zeitschrift, Vol. 106, No. 1-Jan, 1944, pp. 50-62). (16 pages) |
| 2511 | <i>Testing of Aluminum Castings</i> (Die Giesserei, 9.1.42, pp. 7-10). (6 pages) |
| Busemann, A. | <i>On the Increase of Stress in a Plate with Circular Holes and Under Tension</i> (5 pages) |
| 2516 | <i>New Equipment of the Junkers Research Station</i> (Luftfahrtforschung. Vol. 13, No. 2, pp. 67-70, 20.2.1936). (7 pages) |
| 2517 | |
| Fischer, H. | |
| 2518 | |
| Zeerleder, A. | |
| 2519 | |
| Poschl, Th. | |
| 2523 | |
| Roos, H. | |

New Booklets

Interchemical Review, summer issue, 1945, features an article on oriental lacquer work and a discussion of scientific research by Vannevar Bush. Published quarterly by the Research Laboratories of Interchemical Corporation, 432 West 45th Street, New York 19, New York.

North American Philips Company, 100 East 42nd Street, New York 17, New York, has issued a new 8-page booklet entitled *Fine Wire of Special Materials*. It brings the reader up to date on manufacturing methods and problems connected with wires 0.002" to 0.0007" diameter and smaller.

Crystal Research Laboratories, Inc., 29 Allyn Street, Hartford, Connecticut, has published a new catalog describing various types of crystals manufactured by the Laboratories, and their possible usages, both in the communication and industrial fields. 8 pages.

Industrial Research Institute, Inc., 60 East 42nd Street, New York 17, New York, has released for distribution outside the membership of the Institute a monograph on *Organization of Technical Research in Industry*. 16 pages. Copies may be obtained without charge from the Institute.

ELECTRON MICROSCOPE SOCIETY OF AMERICA

THE Electron Microscope Society of America held its third annual meeting November 30–December 1 at the Frick Chemical Laboratory of Princeton University. Over two hundred people attended. In addition to the announced program of papers, abstracts of which are printed below, the following papers were presented in discussion:

H. C. HICKLE AND H. S. DAVIDSON, *Imperial Paper and Color Corporation, Dispersion Methods*

S. F. KERN, *Celanese Corporation of America, Silica Replica Techniques Applied to Fibers*

N. DAVIDSON AND R. F. BAKER, *RCA Laboratories, Shadow Casting Technique in Electron Microscopy*

JAMES HILLIER, *RCA Laboratories, Further Improvements in the Resolving Power of the Electron Microscope*

The following officers were elected:

President, David Harker, General Electric Research Laboratory

Vice President, L. Marton, Stanford University

Secretary-Treasurer, John Turkevich, Princeton University

Director, James Hillier, RCA Laboratories

Director, Ralph W. G. Wyckoff, National Institute of Health

1. Formal Opening of Meeting. JAMES HILLIER, *President*.

2. Welcoming Remarks. HUGH S. TAYLOR, *Dean of Graduate School, Princeton University*.

3. A Brightness Meter for Use on Electron Microscopes. F. W. CUCKOW, *National Physical Laboratory, Teddington, Great Britain*. (Time: 10 min.)—High contrast plates are normally used on electron microscopes and a brightness meter was first used simply as an exposure meter to ensure consistent results. The "reduced beam angle" technique leads to improved picture sharpness and involves visual focusing at a high screen intensity followed by a slight reduction of objective lens current and then exposure at the lower intensity consequent on reduction of beam angle. It has been shown that this procedure can be reduced to a routine giving consistently good results provided that the high and low intensities are both set by reference to the brightness meter and that the objective lens current is reduced by a pre-determined amount.

4. Measurements on Image Distortion in the RCA Type EMU Microscope. ROBERT G. PICARD, *RCA, Victor Division, Camden, New Jersey*. (Time: 15 min.)—A discussion of the lens system of the type EMU microscope. The magnification range and distortion studies and measurements throughout the range will be shown.

5. A Study of Distortion in Electron Microscope Projection Lenses. JAMES HILLIER, *RCA Laboratories, Princeton, New Jersey*. (Time: 15 min.)—The "first-order" theory of distortion in electron microscope projection lenses is presented and is shown to agree with experiment to within the experimental error. It is shown to be possible to obtain a high degree of correction of distortion by the proper

combination of two electron lenses. An experimental double-gap pole-piece for a magnetic projection lens and the distortion patterns obtained with it are also described.

6. New Accessories for Electron Microscopes. PERRY C. SMITH, *RCA Victor Division, Camden, New Jersey*. (Time: 15 min.)—A new design of vacuum unit wherein various materials may be evaporated or sputtered; a new design of vacuum gauge wherein pressures from atmospheric to better than 10^{-4} mm Hg may be measured; an electron spray device for replenishing electrons to specific materials while under electron bombardment in diffraction studies; and a mechanical film spreader for fabricating films for supporting electron microscope specimens will be described.

7. The Preparation of Powdered Materials for Electron Microscopy. MARY C. SCHUSTER AND E. F. FULLAM,* *Interchemical Corporation, New York, New York*. (Time: 15 min.)—The basic problems involved in the preparation of powdered materials for electron microscopy are discussed, with special reference to effecting and maintaining an adequate dispersion of ultimate particles. Various tested methods are described for forming thin films of clear resins, and for mounting dispersed powders on these films. Procedures are also given for dispersing and mounting powders suspended in resin solutions. The properties and applications of various resins and solvents are described, and electron micrographs are shown of several types of mounts.

8. Specimen Mounting Techniques for Inorganic Chemicals. JAMES HILLIER, *RCA Laboratories, Princeton, New Jersey*. (Time: 15 min.)—The mounting of inorganic chemicals for observation in the electron microscope is discussed and a number of techniques are described.

9. A High Speed Microtome for the Electron Microscope. ERNEST F. FULLAM* AND ALBERT E. GESSLER, *Interchemical Corporation, New York*. (Time: 20 min.)—Recognizing the need of a relatively simple method for making thin cross sections suitable for the electron microscope, a high speed microtome was designed after the one built by O'Brien and McKinley. The knife, a piece of razor blade mounted on a wheel rotating at 57,000 r.p.m. moves at speeds up to 1100 feet per second. It is able to make cross sections from (0.1 to 1 micron thick) suitable for stereoscopic electron micrographs of animal tissue, rubber, rayon, Lucite, Nylon, and other materials. A number of sectioning techniques were developed for use with the instrument. One of these employed the use of volatile solid embedding material and their eutectic mixtures to support the sample while being cut and the sections during their collection. Such materials have the advantage that they require no solvent extraction from

* At present at General Electric Company, Schenectady, New York.

the sections with the resulting solvent removal difficulties. A number of cross sections of different materials are shown indicating the possibilities of the apparatus in many fields.

10. The Form of Particles of Nickel Produced by the Carbonyl Process. F. W. CUCKOW, *National Physical Laboratory, Teddington, Middlesex, Great Britain.* (Time: 10 min.)—Electron microscopical examination of (carbonyl) nickel powders consisting of particles a few microns in diameter has revealed two types of powder consisting respectively of (a) a mixture of irregularly shaped and cube shaped particles, each with rough surfaces, and (b) characteristically shaped particles with smooth surfaces. The form of the smooth surfaced particles could not be determined until evidence from the electron microscope was combined with evidence from the light microscope. It was then inferred that the smooth surfaced particles were single crystals in the form of icositetrahedra, probable indices of faces (544).

11. Metal Shadowing of Preparations for Electron Micrography. ROBLEY C. WILLIAMS AND RALPH W. G. WYCKOFF, *University of Michigan, Ann Arbor, Michigan.* (Time: 20 min.)—A technique is described for preparing objects for electron micrography by depositing obliquely onto them a thin film of metal. This procedure, which can be thought of as a form of surface staining, makes evident many details that cannot otherwise be clearly seen and also gives the resulting micrographs a definite three-dimensional quality. Examples will be shown of photographs of such "shadowed" objects as purified viruses and other particles of macromolecular dimensions, replicas of such particles and replicas of surfaces of glass, crystals, and preparations of metallographic interest.

12. A Shadow-Casting Adapter for the Electron Microscope. THOMAS F. ANDERSON, *Johnson Foundation, University of Pennsylvania, Philadelphia, Pennsylvania.* (Time: 10 min.)—A device fitting on the specimen chamber door seat is described which permits the evaporation of metal onto the specimen in the high vacuum of the electron microscope itself. The specimen can be observed during the shadow-casting process to insure that proper thicknesses of metal are obtained. Results are illustrated with bacteria and bacterial viruses.

13. Metal Shadowing with Virus. D. GORDON SHARP, *School of Medicine, Duke University, Durham, North Carolina.* (Time: 15 min.)—Electron pictures showing marked three-dimensional properties have been obtained with virus by the method of metal shadowing.¹ Metal molecules deposited in vacuum by evaporation on the surface of electron microscope preparations are made to cast shadows of the virus particles. These preparations are then photographed in the electron microscope by transmitted electrons in the usual way. Examples of the work done in this way will be shown together with conventional electron micrographs of the same materials.

¹ R. C. Williams, and R. W. G. Wyckoff, *Proc. Soc. Exp. Biol. and Med.* 58, 265 (1945).

14. Report of the Committee on Resolving Power. (Time: 45 min.)

15. Recent Developments in the Field of Electron Microscopy (Invited paper). RALPH W. G. WYCKOFF, *National Institute of Health, Bethesda, Maryland.*

16. A Study of Oxide Films Formed on Metals and Alloys by Electron Microscopy and Diffraction. R. T. PHELPS AND EARL A. GULBRANSEN, *Westinghouse Research Laboratories, Pittsburgh, Pennsylvania.* (Time: 20 min.)—Highly polished samples of metals and alloys are oxidized at controlled conditions of temperature and time in a furnace attached to a diffraction camera. Diffraction patterns by the reflection technique are obtained on the samples before and after oxidation at the temperature of oxidation, also at room temperature after cooling. The thin oxide films, 100 to 400 Å thick, are stripped from the oxidized metal by chemical or electrochemical means for examination by transmission electron microscopy and diffraction. Electron micrographs reveal the following information concerning the oxide film: (1) particle size, (2) particle size distribution, (3) particle shape, (4) uniformity of film thickness, and (5) type of micrograph. Diffraction patterns are used to identify the composition of the films and to indicate approximate particle sizes. Complete agreement between reflection and transmission diffraction patterns is not expected in the films containing more than one metal or more than one oxidation state of a metal. Inherently the two methods sample the film differently.

17. An Electron Diffraction Study of Oxide Film Formation on Metals and Alloys. EARL A. GULBRANSEN AND J. W. HICKMAN, *Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania.* (Time: 20 min.)—The reflection method is applied to the study of the oxidation reaction of metals with air or oxygen atmospheres. The experiments are carried out "in situ" over the temperature range of 200 to 700°C and over a wide time range. Existence charts of the oxides formed are plotted as a function of time and temperature of the oxidation reaction. These charts are used to correlate the structure with the protective nature of the oxide film. The factors influencing the formation of a particular oxide are briefly discussed.

18. Investigation of Etched Metallic Surfaces by Electron Diffraction and the Electron Microscope. R. D. HEIDENREICH,* L. STURKEY AND H. L. WOODS, *The Dow Chemical Company, Midland, Michigan.* (Time: 20 min.)—A method of rinsing metal surfaces immediately after etching is described which produces surfaces that are "clean" as far as electron diffraction is concerned. Examples of the method are presented including magnesium alloys, steel and copper alloys. The combined technique of electron microscopy and electron diffraction proves to be very sensitive to minute amounts of secondary phases and

* Now at Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

impurities in etched metals. The decomposition of martensite at 200°C is shown to produce a colloidal mixture the slow etching component of which is Fe_3N . At 400°C the slow etching component is the usual Fe_3C . The nitride formed at 200°C can be converted to the carbide by further heating at 350°C. The position of these results relative to the earlier x-ray diffraction work is discussed. In the case of an aluminum-copper alloy, the usual etchants are not satisfactory due to the re-deposition of the copper on the etched surface as polycrystalline metal. Pure aluminum can be etched to produce clean surfaces. Difficulty with small amounts of impurities in copper alloys is illustrated by β -brass containing 0.0005–0.001 percent Ag. When etched in an acetic-nitric acid reagent, the silver impurity is dissolved and then precipitated onto the etched surface in the form of spherical particles about 1000Å in diameter. Electron micrographs and electron diffraction reflection patterns for beryllium copper are shown although a systematic study of this system has not been completed. The possibilities for research on metals using these techniques are discussed not only from the standpoint of metallurgical structures but from that of studying the formation of surface films in various atmospheres.

19. The Application of the Electron Microscope to the Study of Metals (Invited Paper). DAVID HARKER, *Research Laboratory, General Electric Company, Schenectady, New York.* (Time: 60 min.)

20. An Application of the Electron Microscope to the Study of Animal Cells (Invited Paper). KEITH R. PORTER, *The Rockefeller Institute for Medical Research, New York, New York.* (Time: 60 min.)

21. Micrographs of Insect Wings. E. F. BURTON, *Department of Physics, University of Toronto, Toronto, Ontario, Canada.* (Time: 10 min.)

22. An Investigation of Transverse Striations and Myosin Filaments in Muscle. C. E. HALL, M. A. JAKUS AND F. O. SCHMITT, *Massachusetts Institute of Technology, Cambridge, Massachusetts.* (Time: 20 min.)—Electron microscope specimens of striated muscle were obtained by fragmenting fixed muscles in a blender. This treatment tends to split the fibre bundles into their constituent fibrils (myofibrils) along natural cleavage boundaries. Electron microscope examination of such fibrils permits a more detailed description of the transverse striations previously observed with the light microscope and provides additional information on other structures below light microscope resolution. The fibrils consist chiefly of parallel myosin filaments ranging in width from 50 to 250Å. The filaments

extend continuously through the anisotropic and isotropic bands and pursue a straight course in both extended and contracted states. Electron micrographs were also made of myosin filaments obtained by extracting fresh muscle in weakly alkaline salt solutions after the method of Edsall. Statistical measurements were made of lengths and widths of filaments from rabbit, frog, lobster, scallop, and clam. Lengths vary from 500 to 20,000Å and show significant variations from one animal form to the next. Widths range from 50 to 250Å and show relatively little variation between forms.

23. The Structure of the Trichocysts of Paramecium. MARIE A. JAKUS, *Massachusetts Institute of Technology, Cambridge, Massachusetts.* (Time: 15 min.)—The minute size of trichocysts makes them a suitable object for electron microscope investigation. Extruded trichocysts can be readily obtained by stimulating electrically a natant paramecium on an electron microscope specimen grid. Each trichocyst is found to consist of a dense pointed tip and an elongated shaft which appears cross-striated in electron micrographs. It is suggested that the shaft consists chiefly of thin cylindrical membrane formed by the close packing of longitudinal fibrils, in phase with respect to a periodic axial pattern. The structure and chemical properties of this membrane are discussed.

24. Electron Microscope Studies of the Structure of Chloroplasts, Chromoplasts, Leucoplasts Showing their Relationship to "Cellulose Particles," Carotin "Crystals" and Colloidal Carbon. EDITH A. ROBERTS, *Vassar College, Poughkeepsie, New York.* (Time: 10 min.)—The photographs taken with the electron microscope would indicate that the granules of which the chloroplasts, chromoplasts, and leucoplasts are composed are the same as the "cellulose particles" of which cell walls are built, the bodies of which carotin "crystals" are composed and the same as those found in colloidal carbon. All these granules are made up of similar primary (0.5 μ); secondary (0.25 μ); tertiary (0.15 μ); quarternary (0.05 μ); quintary (0.02 μ); and senary (0.01 μ) units.

25. Electron Microscope Studies of the Structure of Starch, Showing its Relationship to Plastid Structure. MILDRED D. SOUTHWICK, *Vassar College, Poughkeepsie, New York.* (Time: 10 min.)—Studies are made of the structural organization of starch grains from several genera of plants using the electron microscope. These starch grains are taken from varied storage organs—seed, fruit, root, or tuber. The structure of the starch grain is compared with that of the chloroplast and the starch bodies which appear in the chloroplasts of the same plant.

Journal of Applied Physics

Volume 17, Number 2

February, 1946

Problems in Non-Elastic Deformation of Metals*

BY CLARENCE ZENER** AND J. H. HOLLOMON

Research Division, Watertown Arsenal, Watertown, Massachusetts

In this article an attempt is made to examine the goals of current and of possible future types of research concerned with the non-elastic deformation of metals. It is found that an understanding of non-elastic deformation would be enhanced by research upon the following topics: 1. The anelasticity associated with the viscous behavior of grain boundaries and the at least temporary viscous behavior of slip bands. 2. The mechanics of the initiation and growth of twin bands. 3. The mechanics of the initiation and growth of slip bands, including the drop in

resistance to deformation which accompanies the initial slip bands. 4. The mechanics of the segregation of solute atoms in solid solution, such as of carbon and nitrogen in iron. 5. The conditions under which strain hardening is not removed by recovery or by recrystallization, and hence under which a mechanical equation of state exists, i.e., under which a relation exists between strain rate, strain, stress, and temperature. 6. Variation of the heat of activation for plastic strain rate upon stress and upon the microstructure. 7. Anisotropy introduced by deformation.

IN all fields of investigation it is advisable now and then to assess the potentialities of the various paths along which further research may profitably proceed. Such an assessment should be particularly valuable in the fields of plastic deformation and fracture of metals, because of the large number of variables involved in the types of experiments which may be performed and in the types of materials to be investigated. In this and an accompanying report, attempts are made to examine the goals of current and possible future types of research concerned with the non-elastic deformation and fracture of metals.

I. INTRODUCTION

In an idealized metal the stress-strain relation consists of two parts, elastic and plastic. The

essential feature of the first part is that of reversibility, i.e., the uniqueness of the relation. This relation is usually, but is not necessarily, regarded as linear. The essential feature of the second part is the presence of a permanent set. Thus in Fig. 1 the path ABB' is traced in loading, $B'A'$ is traced in unloading, while the permanent set AA' remains. The slope of the stress-strain relation in the plastic region is usually, but is not necessarily, positive.

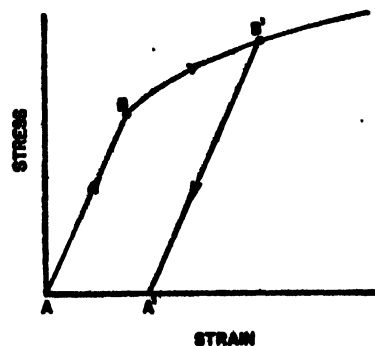


FIG. 1. Idealized stress-strain relation.

* The statements or opinions in this article are those of the authors and do not necessarily express the views of the Ordnance Department.

** Now at the Institute for the Study of Metals, University of Chicago, Chicago, Illinois.

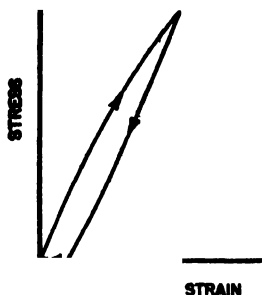


FIG. 2. Typical stress-strain relation of an actual metal in the pre-plastic region.

No real metal possesses an ideally elastic region. Even though a definite yield stress may exist below which no permanent set occurs, the strain is a function of the past history of the stress as well as of its instantaneous value. A stress-strain curve in the pre-plastic region may thus be like that represented in Fig. 2; the recovery takes place gradually after the removal of the load. If the observations cease before the hysteresis loop is essentially complete, the observer will deduce that the metal has acquired a permanent set, and has thus behaved plastically. That property of a solid in virtue of which stress and strain are not uniquely related in the pre-plastic range has been called anelasticity. An appreciation of anelastic phenomena is therefore essential to a complete understanding of plastic deformation.

The plastic part of the stress-strain curve in Fig. 1 is idealized in two respects. Firstly, in the case of alloys the curve as observed is often jagged, consisting of "serrations." Secondly, the curve for a real metal approximates that in Fig. 1 only if certain parameters are maintained constant during the test, such as strain rate and temperature, parameters which under many conditions of strain do not remain constant.

No intrinsic difficulties prevent one from deducing the elastic part of the idealized stress-strain curve under any stress pattern from the simple tensile test, provided Poisson's ratio is also measured. Similarly, one may predict the plastic portion of the idealized stress-strain curve for an arbitrary stress pattern from data obtained in simple tension, upon taking Poisson's ratio equal to $\frac{1}{2}$. Such predictions are not however as successful as in the elastic case.

From the above brief introduction it is clear that the non-elastic behavior of metals cannot be described in simple terms except under very special conditions. In order to understand the multitudinous ways a metal may respond to the application of stress, it will be necessary first to understand the details of deformation on a microscopic scale.

The physical origins of anelasticity have recently been reviewed by one of the authors.¹ The origins are of two general types. The first type is associated with some sort of diffusion induced by applied stresses, e.g., thermal diffusion, atomic diffusion, magnetic flux diffusion. The second type is associated with isolated regions which behave in a viscous manner. Whatever the origin of the anelasticity, Boltzmann² has shown how one may predict the behavior of the speci-

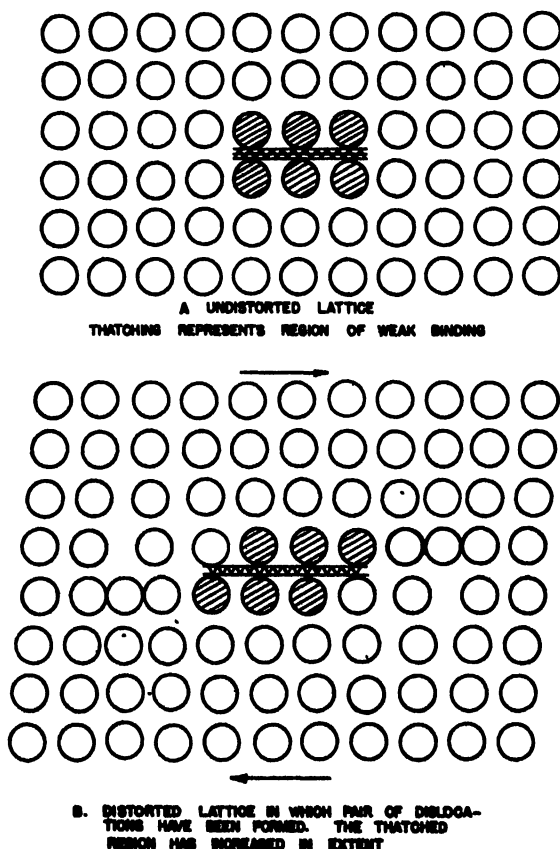


FIG. 3. Illustration of initiation of dislocations.

¹ C. Zener, "Anelasticity of metals," Trans. A. I. M. E. (in press).

² L. Boltzmann, "Zur Theorie der elastische Nachwirkung," Ann. d. Physik 7, 624 (1876); also Sitz. Wien Akad. [2] 70, 275 (1874).

men under any conditions of loading from observations made upon the response when a fixed load is suddenly applied.

A detailed description of one microscopic mechanism for plastic deformation in crystal lattices was first given by Orowan³ based upon suggestions of Polanyi. According to this description, plastic deformation takes place in two stages. In the first stage lattice distortions of a particular type are produced in the vicinity of a region of low shear resistance. These distortions are known as dislocations. Their mode of generation is illustrated in Fig. 3. In the second stage the dislocations are propagated through the lattice. This propagation is hindered by any type of lattice disturbance, such as other dislocations (which hindrance gives rise to strain hardening), precipitates, grain boundaries, etc.

A second possible mechanism of plastic deformation is by the formation of twins. No satisfactory dynamic theory has yet been advanced as to how such twin formation occurs.

If these four mechanisms, diffusion, viscous flow in isolated regions, dislocations, and twinning, are the only mechanisms for non-elastic behavior of metals, it must be possible to interpret all observed types of non-elastic behavior, and the influence thereon of microstructure, in terms of one or other of these mechanisms.

II. ANELASTICITY

Viscous flow in isolated regions may, on the other hand, cause deviations of more than 100 percent from perfect elasticity. Viscous flow may be expected in regions where the atoms are arranged in an amorphous, or non-crystalline, manner. Thus the atoms on the two sides of grain boundaries cannot be aligned with one another, and so such boundaries might be expected to behave in a viscous manner. The profound effects of viscous slip upon the stress-strain relation may be seen from Fig. 4. This figure has been constructed from the recent work⁴ of W. A. West of this laboratory. One practical problem of great importance in which viscous slip at grain boundaries may play a dominant

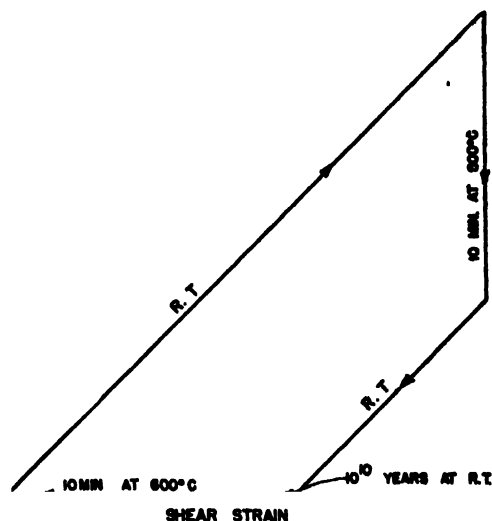


FIG. 4. Anelastic behavior of an iron wire due to viscous flow at grain boundaries, after experiments of West (reference 4).

role is stress relaxation at elevated temperatures. Thus in the example of Fig. 4, 60 percent of the stress was relaxed by a recoverable anelastic deformation associated with grain boundary slip, not more than 6 percent by plastic deformation. Another problem of even greater practical importance in which viscous slip at grain boundaries may play a dominant role is creep at elevated temperatures when close tolerances must be maintained. Thus with the iron specimen from which the data of Fig. 4 were taken, ten minutes at 500°C would more than double the strain under conditions of constant stress. At least 95 percent of this additional strain would be of the recoverable anelastic type associated with grain boundary slip.

Rosenhain⁵ recognized the importance of viscous grain boundary slip in the 1910's, and performed many experiments demonstrating the existence of such slip. In spite of the above-mentioned practical importance of this phenomenon, no experiments have been performed during the intervening 30 years to measure in any metal the relation between rate of slip across the grain boundary and the stress which causes such slip. This complete absence of experimental data is

³ E. Orowan, "Zur Kristallplastizität," Zeits. f. Physik 89, 605, 614, 634 (1934).

⁴ W. A. West, "Elastic after-effects in iron wires from 20°C to 550°C," communicated to A. I. M. E.

⁵ W. Rosenhain and D. Ewen, "Intercrystalline cohesion in metals," J. Inst. Metals [2], 8, 149 (1912); also *Introduction to Physical Metallurgy* (Constable and Company, London, 1915), pp. 255-264.

due to the fact that nearly all physical testing of metals, especially in this country, has been, and is still being, done with an entire disregard of the microstructure of the metal. This procedure is a reflection of the lack of faith in the ability of physical science to predict the gross mechanical behavior of metals in terms of the mechanical behavior of the individual constituents, a faith to which the present authors adhere.

Until an atomistic interpretation of all non-elastic effects is forthcoming, it will be useful to have certain phenomenological concepts regarding the mechanical behavior of microscopic elements larger than atomic dimensions. Such a concept is, for example, that of viscous grain boundaries. Another concept which may prove of value is that of viscous slip bands, a concept which may be studied by anelastic measurements.

It was recognized as early as 1900 that plastic deformation was confined to slip bands,⁶ that such deformation introduced anelastic phenomena, such as mechanical hysteresis, and that this anelasticity was removed⁷ by annealing at temperatures as low as 100°C. Rosenhain⁸ pointed out in 1905 that these effects of plastic deformation indicated that the material in freshly formed slip bands behaved as if it were disorganized, i.e., in a viscous manner, and that upon aging this material gradually assumed the order of the surrounding matrix. During the intervening 40 years no experiments have been performed for the purpose of further elucidating the mechanical behavior of freshly formed slip bands. Many observations have been made upon special aspects of the anelasticity introduced by plastic deformation. These have been reviewed in reference 1. It is not to be expected, however, that such observations can contribute to an understanding of the mechanics of deformation until they are interpreted in terms of the mechanical properties of the constituent slip bands.

⁶ J. A. Ewing and W. Rosenhain, "Experiments in micrometallurgy—effects of strain," *Phil. Trans. Roy. Soc.* 193, A353 (1900).

⁷ J. Muir, "Recovery of iron from overstrain," *Phil. Trans. Roy. Soc.* 193, A1 (1900).

⁸ W. Rosenhain, "Deformation and fracture in iron and steel," *J. Iron and Steel* [2] 70, 189 (1906).

III. MECHANISM OF PLASTIC DEFORMATION

Plastic deformation in metals, that is, permanent deformation, may occur by two distinct mechanisms: deformation by twinning and deformation by slip. In any particular case deformation may proceed by any combination of these two types.

Two adjacent parts of a crystal separated by a plane are said to be twins when their crystallographic axes are different but are so related that each part is the mirror image of the other part with respect to their common plane. The crystallographic relations between known twins are given by Schmid and Boas⁹ and by Barrett.¹⁰ Twinning is deformation through the formation of twins. The final configuration may be thought of as having been attained by a homogeneous simple shear parallel to the common plane. Twinning is not associated with any lattice distortion.

Deformation by slip is said to occur whenever strain is confined to narrow bands which are not twins of the adjacent region. The orientation of the slip bands, as well as the direction of slip, are observed to have certain crystallographic relations with the parent lattice. These relations are given by Schmid and Boas¹¹ and by Barrett.¹² The material within the slip band has in general the same orientation as the parent lattice, but suffers considerable fluctuation from region to region.

Temperature and rate of strain are the primary factors which determine in what manner a crystalline material will deform. As the temperature increases, or the rate of deformation decreases, the resistance to deformation by slip decreases rapidly, resistance to deformation by twinning decreases less rapidly, if indeed it decreases at all. Stress gradient is another factor which has some influence in determining the manner in which deformation will occur. A macroscopic deformation may be thought of as the superposition of many small deformations, the smallest possible deformation being called an

⁹ E. Schmid and W. Boas, *Kristallplastizität* (Verlagsgesellschaft Julius Springer, Berlin, 1935), p. 100.

¹⁰ C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 312.

¹¹ Reference 10, p. 90.

¹² Reference 10, p. 289.

elementary deformation. Each elementary deformation is essentially confined to a limited region. To a first approximation only the average of the stress throughout the region is important in determining whether an elementary deformation will occur. In deformation by twinning and by slip the lateral dimensions of the region are comparable to the crystal size. Whether stress localization favors slip over twinning, or *vice versa*, is not at present known. Composition is another factor which has some influence over the type of deformation which occurs. Thus the addition of silicon to iron raises the resistance to deformation by slip relative to the resistance to deformation by twinning.¹³ Whether this effect of silicon is common to all solid solutions is not known at present, nor is the relative effect of insoluble precipitates known.

A. Twinning

As was probably first observed by Chalmers¹⁴ in his experiments upon tin, deformation by twinning is a discontinuous process on a macroscopic scale, in other words, the region of an elementary deformation is of macroscopic dimensions. It would be anticipated that under conditions of a uniform stress a critical shear stress would exist for the initiation of a twin. Actually, however, the conditions for the initiation of a twin appear to be so structure sensitive that in no case has a reproducible resolved critical shear stress for twin formation been obtained in work upon single crystals. Unlike deformation by slip, deformation by twinning may start in a local region of high stress concentration and propagate for a long distance through regions which were under no stress prior to deformation. (See Fig. 47 of Schmid and Boas.) One might suspect that this ability of twins to be set off by local stress concentrations would render the critical resolved shearing stress for twin initiation sensitive to surface irregularities. This possible effect of surface irregularities has not been investigated.

Once a twin band has been formed, the band may grow continuously by the propagation of the

twin interfaces. Thus the twin bands of recrystallized alpha-brass grow broader as the grains grow. No studies have been made upon the dependence of the velocity of these interfaces with the various parameters, such as with stress and temperature.

Phase transformations by lattice shear have many features in common with twinning. Examples are the formation of martensite in steel and of martensite-like structures in beta-brass. Any progress in understanding the mechanics of one phenomenon will aid in the understanding of the other.

B. Slipping

As a first approximation, deformation by slip may be regarded merely as the slipping of one set of planes upon another. This picture is sufficiently accurate to predict the planes and directions of slip. If one plane be regarded as uniformly slipping over another plane, the potential energy of the system will undergo a periodic fluctuation. The amplitude of the fluctuation will be least for planes of closest packing, both because the atoms are closest together in these planes and because these planes are the furthest apart. Similarly, the fluctuation will be least along those directions in which the atoms are most closely packed. Thus the picture of simple slip interprets satisfactorily the observation that the planes of easiest slip are the planes of closest packing, and that the direction of slip is the direction along which the atoms are closest.

The concept of a homogeneous slip of one plane of atoms over another cannot however be correct. Such slip would require for an unalloyed crystal a shearing stress several orders of magnitude larger than is actually observed. Polanyi¹⁵ was apparently the first to realize how this difficulty could be resolved through the introduction of a certain type of lattice distortion, known in English as "dislocations." These dislocations will move through the lattice, and thereby result in a net macroscopic strain, under a considerably smaller stress than would be necessary to cause two planes to slip over one another in a homogeneous manner. The full effectiveness of dis-

¹³ C. S. Barrett, G. Ansel and R. F. Mehl, "Slip, twinning and cleavage in iron and silicon ferrite," *Trans. Am. Soc. Metals* **25**, 702 (1937).

¹⁴ B. Chalmers, "Twinning of single crystals of tin," *Proc. Phys. Soc. (London)* **47**, 733 (1935).

¹⁵ M. Polanyi, "Über eine Arte Gitterstörung, die einen Kristall plastisch machen konnte," *Zeits. f. Physik* **89**, 660 (1934).

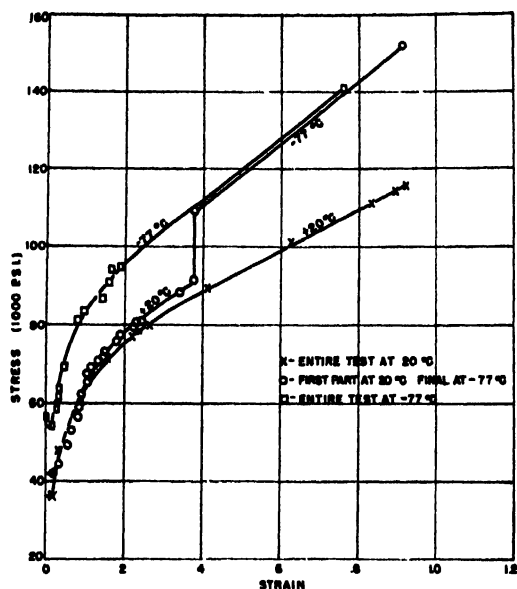


FIG. 5. Example of validity of mechanical equation of state for a plain carbon steel at low temperatures. (Tensile tests.)

locations in reducing the resistance to slip can be appreciated only by a consideration of the thermal energy of the lattice. The fluctuations in the thermal energy will, given time, give rise to a unit slip when the stress by itself is not sufficient to cause slip. Such fluctuations in thermal energy would not aid one plane of atoms to slip homogeneously over another plane, for the probability of a large enough fluctuation would be too small. Thus suppose E is the directed kinetic energy which the system must have in order to aid the stress in producing a unit slip. The probability that at any given time the system possesses directed kinetic energy equal or greater than this amount varies as $\exp(-E/kT)$. The thermal fluctuations can therefore be of significance only when E is not many orders of magnitude greater than kT , therefore only when a unit slip occurs in a comparatively small volume, as in the case of dislocations.

If the only effect of temperature upon plastic deformation is through the introduction of energy fluctuations, it must follow that at a constant stress the strain rate will be higher the higher the temperature, or, conversely, the stress necessary to produce a given strain rate will be lower the higher the temperature. At sufficiently high temperatures additional effects occur, such as

grain growth, spheroidization of precipitates, etc., and, if the prior deformation is sufficiently large, recovery with or without recrystallization. At lower temperatures when such additional effects do not occur it might be expected, as was pointed out by Ludwik,¹⁶ that a mechanical equation of state exists which relates stress s , strain ϵ , strain rate $\dot{\epsilon}$, and temperature T . Such an equation may be written symbolically as

$$S = S(\epsilon, \dot{\epsilon}, T), \quad (1)$$

or as

$$\dot{\epsilon} = \dot{\epsilon}(\epsilon, S, T). \quad (2)$$

The existence of a mechanical equation of state would imply that the stress necessary to produce a given strain rate at a given temperature depends only upon the strain and not upon the temperature or strain rate at which the prior strain occurred. According to Taylor's concept,¹⁷ a discontinuous change in temperature would not produce a discontinuous change in stress, only a discontinuous change in rate of increase of stress with strain. In order to determine whether a mechanical equation of state exists, experiments have been performed in this laboratory in which the temperature was suddenly changed during the determination of a stress-strain curve. The results, presented as Fig. 5, show that the stress is independent of the temperature at which prior deformation occurred, and therefore indicate that a mechanical equation of state does in fact exist (at least at low temperature), and therefore that, at the temperature of the test, the only influence of temperature is through the introduction of fluctuations in thermal energy.

Both deformation by twinning and by slip are processes characteristic of crystals. Hence any kind of disturbance of the crystalline structure would be expected to raise the resistance to deformation by twinning and by slip. This is indeed the case. Such disturbance may be due to prior plastic deformation, to grain boundaries, to soluble impurities, to insoluble impurities, etc. The increase in resistance to plastic deformation by prior deformation has been investigated theoretically by Taylor¹⁷ using the concept of

¹⁶ P. Ludwik, *Elemente der technologischen Mechanik* (Verlagsbuchhandlung Julius Springer, Berlin, 1909).

¹⁷ G. I. Taylor, "Mechanism of plastic deformation of crystals, Part One," *Proc. Roy. Soc.* 145A, 362 (1934).

dislocations. Although Taylor's theory must be regarded as a milestone on the road leading to understanding strain hardening, it was based upon assumptions too naive to allow it to pass the test every physical theory should meet, the prediction of new phenomena. No more recent analysis has eliminated the undesirable assumptions, e.g., the assumptions of impenetrable barriers and of parallel dislocations of indefinite length. The increase in resistance to plastic deformation caused by grain boundaries has likewise been investigated theoretically by Taylor.¹⁸ While his analysis reproduces precisely the difference between the stress-strain curves of aluminum single crystals and of aluminum polycrystalline specimens, it does not provide an explanation of the frequently observed variation^{19,20} of resistance to deformation with grain size in polycrystalline material. Taylor assumed in essence that the effect of the restraints imposed by adjacent grains extended uniformly through the grains, while the above-mentioned observations indicate that the effect of the restraints is confined more or less to the vicinity of the grain boundaries. No modifications have as yet been made of Taylor's theory to take account of the localization of the effects of the restraints. A correlation has been found by Gensamer²¹ of the increase in resistance to plastic deformation introduced by solute atoms and the difference in size of the solute and solvent atoms. A correlation has been found by Gensamer and his collaborators²² of the increase in resistance to plastic deformation introduced by insoluble precipitates and the mean free path between these precipitates. As yet no satisfactory theory has interpreted these two types of correlations.

Although, on the whole, deformation raises the resistance with respect to further deformation,

the same cannot be said of each element of deformation. A slight deformation may in fact momentarily lower the resistance with respect to further deformation. Thus a lower stress is required to propagate a slip band than to initiate it. Further, in many cases a lower stress is required to initiate a slip band adjacent to one already formed than to initiate the original band. This momentary lowering of resistance to further deformation by a small deformation gives rise to a diversity of phenomena. Thus when a constant stress is applied to the specimen, elongation may occur in a step-wise fashion. On the other hand, when the over-all strain rate is maintained constant, the stress may undergo a series of discontinuous jumps. Whether each jump is associated with a single slip line, or with a series as in the case of Duralumin, depends to a great extent upon the elasticity of the system. Thus the elastic strain in the specimen and test machine may be only sufficient to produce a single slip band, or it may be sufficient to produce a whole series of slip bands. In spite of a large number of experiments in which discontinuous yielding has been observed and studied,* no satisfactory theory has been presented which accounts for the momentary drop in resistance upon yielding, and the attendant phenomena. The ideas advanced by Orowan³ concerning the origin of dislocations may form the basis for such a theory. According to these ideas, the dislocations originate in some weak region of the lattice, represented by the hatched area in Fig. 3. As may be seen in illustration *b* of this figure,

* *Drop in Load on Yielding in Mild Steel*: J. H. Wicksteed, Proc. Inst. Mech. Eng., p. 26 (1886). *Serrations in Mild Steel*: A. Le Chatelier, Revue de Metallurgie 6, 914 (1909). *Discontinuous Yielding in Copper*: Andrade, Proc. Roy. Soc. 84, 1 (1916). *Serrations in Duralumin*: A. Portevin and F. Le Chatelier, Trans. A. S. Steel Treat. 5, 457 (1924); R. Anderson, Proc. A. S. T. M. [2] 26, 349 (1926); N. Dawidenkow, Zeits. f. Physik 61, 46 (1930). *Discontinuous Yielding in Alpha Brass*: G. Sachs and H. Shoji, Zeits. f. Physik 45, 776 (1921); W. Köster, Zeits. f. Metallkunde 19, 309 (1927); M. Classen-Nekludowa, J. d. Russ. Phys. Ges. 60 (1928). *Discontinuous Yielding in Rock Salt*: M. Classen-Nekludowa, J. d. Russ. Phys. Ges. 59 (1927); Obreimow and Schubnikoff, Zeits. f. Physik 41, 907 (1927); A. Joffé, *The Physics of Crystals* (McGraw-Hill Book Company, Inc., New York, 1928); M. Classen-Nekludowa, Zeits. f. Physik 55, 555 (1929); N. Dawidenkow, *ibid.* 61, 46 (1930). *Discontinuous Yielding in Zinc and Cadmium Crystals*: E. Schmid and Valouch, Zeits. f. Physik 75, 513 (1932); R. Becker and E. Orowan, *ibid.* 79, 566 (1932); E. Orowan, *ibid.* 89, 605 (1934).

¹⁸ G. I. Taylor, "Plastic strain in metals," J. Inst. Metals 62, 307 (1938).

¹⁹ C. A. Edwards and L. B. Pfeil, "Tensile properties of single iron crystals and the influence of crystal size upon the tensile properties of iron," J. Iron and Steel Inst. 112, 79 (1925).

²⁰ W. A. Wood, "X-ray studies of grain size in steels of different hardness values," Phil. Mag. 10, 1073 (1930).

²¹ M. Gensamer and C. E. Lacy, Trans. A. S. M. E. 32, 88 (1944).

²² M. Gensamer, E. B. Pearsall, W. S. Pellini, and J. R. Low, "Tensile properties of pearlite, bainite and spheroidite," Trans. Am. Soc. Metals 30, 983 (1942).

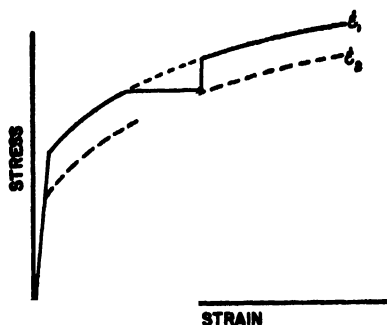


FIG. 6. Example of stress-strain relation when stress is temporarily maintained constant.

after one pair of dislocations has been formed and has propagated away from its origin, the weak region has become extended. A smaller stress will now be required to form a new pair of dislocations. The formation of dislocations thus becomes a cataclysmic phenomenon.

Discontinuous yielding is especially marked in mild steel and in dural. In each case the phenomenon may be related to a state of supersaturation, in iron with carbon and/or nitrogen, in aluminum with copper or zinc. Thus the discontinuous phenomenon disappears when all the carbon and nitrogen are removed from the iron, and becomes much less marked in aluminum after aging. It looks as if in these cases the phenomenon is associated with the presence of flake-like regions in which the solute atoms have either become especially highly concentrated, or have precipitated. Such flake-like regions appear commonly to precede visible precipitation, and presumably arise from the circumstance that such a distribution decreases the strain energy of the system.²³ These flakes will act as restraints with respect to the passage of dislocations. Once a dislocation has crossed a flake, thereby separating it into two halves, the flake offers no further resistance to dislocations along the same plane. Many problems relating to this phenomenon remain to be solved. For example, the phenomenon does not occur in single crystals of iron, but it does in single crystals of aluminum alloys. Again, the theory of the formation of concentration flakes has not yet been developed.

²³ N. F. Mott and F. R. N. Nabarro, *Proc. Phys. Soc. (London)* **52**, 86 (1940).

IV. GROSS FEATURES OF PLASTIC DEFORMATION

A. Mechanical Equation of State

The assumption of a mechanical equation of state leads to many interesting qualitative predictions. As an example, suppose a specimen is strained isothermally at a constant rate until a strain ϵ_0 is reached; the stress is then maintained constant for a long interval of time, and straining is then resumed at the initial rate. A plot of the observed stress-strain curve obtained under such conditions is given in Fig. 6. While the stress is maintained constant the strain is increasing at a rate given by Eq. (2), and hence the stress necessary to resume straining at the original rate is also increasing in accordance with Eq. (1). Stress-strain curves of this type have in fact been observed by Hanson and Wheeler.²⁴ As a second example, we shall consider that the load rather than the stress is maintained constant, an example which has been discussed in detail by Ludwik. As may be seen from Fig. 7, the sudden application of a load will give rise to an initially large strain rate, a strain rate which gradually decreases, reaches a minimum at a strain E_0 , and thereafter increases. The diagram of strain *vs.* time will therefore look as depicted in Fig. 7, the inflection in the curve corresponding to the initiation of necking.

The strain *vs.* time relation depicted in Fig. 7 is precisely of the type encountered in creep tests. In spite of the fact that Ludwik's work was published over 40 years ago, the universal interpretation of the constant creep rate portion of the creep curves is that it represents a balance between the rate of strain-hardening and of softening by recovery or by recrystallization. A decision as to which interpretation of creep curves is correct could be reached by experiments in which the temperature was varied during the test, e.g., by straining 1 percent at room temperature, heating to the temperature at which creep tests are made, cooling back to room temperature, and finally determining if the hardening introduced by the initial small strain has been reduced. It is to be particularly em-

²⁴ D. Hanson and M. A. Wheeler, "Deformation of metals under prolonged loading. Part I. Flow and fracture of aluminum," *J. Inst. Metals* **45**, 1 (1931).

phasized that the determination of a stress function $S_T(\epsilon, \dot{\epsilon})$ at a series of temperatures by experiments performed at constant strain rates, as discussed by Nadai,²⁵ does not establish the existence of a mechanical equation of state. In order that such a function represent a mechanical equation of state it is necessary for the additional observation to be made that the stress required for further deformation is not lowered with time as the strain is maintained constant.

Two approaches have been followed in an attempt to establish the form of the mechanical equation of state for metals. Since neither approach has been entirely successful, it will be of profit to examine them in some detail in order to discern along what path further research may be successful.

The first approach was indicated by Becker²⁶ in 1926, and was further elucidated by Orowan³ in 1936. Here attention is focused upon the shearing stress which an elementary region must possess in order that it undergo plastic deformation. Suppose S_0 is the net shearing stress necessary for deformation, and S is the macroscopic shearing stress. Then thermal fluctuations must provide the additional stress ($S_0 - S$). If V is the minimum volume in which the stress S_0 must exist in order that deformation occur, then the strain rate will be given by

$$\dot{\epsilon} = f_0 \exp \left[-V(S_0 - S)^2 / 2GkT \right]. \quad (3)$$

The second factor is the probability that a given region will at any time have a thermal shear stress of magnitude greater than $S_0 - S$. The first factor f_0 may be thought of as being composed of the following factors

$$f_0 = n\epsilon_0 / \tau, \quad (4)$$

where n is the number of effective regions of volume V per unit volume, ϵ_0 is the mean strain induced by a unit plastic deformation in one region per unit volume, and τ is the mean time an elementary region must possess the shearing stress S_0 before a shear deformation occurs therein. In his discussion of Becker's mechanical equation of state, Orowan has pointed out that

²⁵ A. Nadai, *Plasticity* (McGraw-Hill Book Company, Inc., New York, 1931), Chap. 39.

²⁶ R. Becker, "Über Plastizität, Verfestigung und Rekristallisation," *Zeits. f. tech. Physik* 7, 547 (1926).

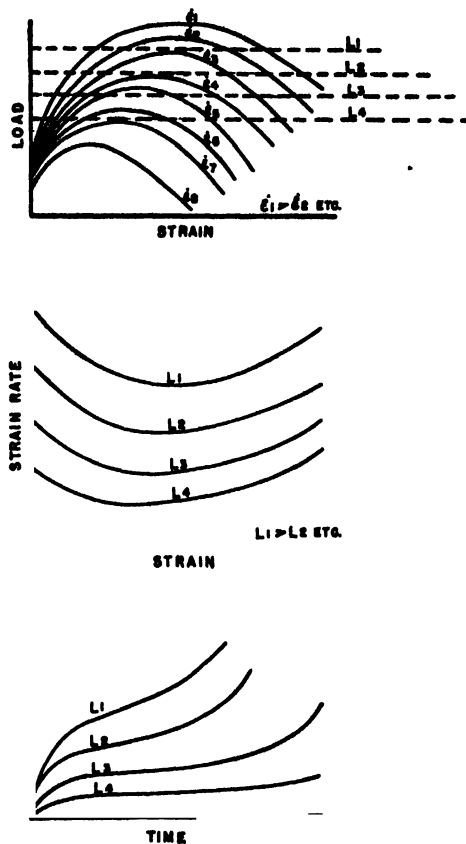


FIG. 7. Derivation of strain-time curves according to concept of mechanical equation of state.

S_0 is of the same order of magnitude as the macroscopic shear stress needed to induce deformation at room temperature. From this observation Orowan concluded that deformation was confined, at any one instant, to locales of high stress concentration.

The second approach was initiated by Condon²⁷ in 1938, and expanded by Kauzmann²⁸ and by Dushman²⁹ and collaborators. Here attention is focused upon the activation energy necessary for an elementary slip process. If Q_0 is the activation energy in the absence of an externally applied stress, when a stress S is applied, the activation energy is assumed to be lowered by an amount proportional to S , thus

$$Q(S) = Q_0 - VS.$$

²⁷ E. U. Condon, *Trans. A. I. M. E.* 131, 410 (1938).
²⁸ W. Kauzmann, "Flow of metals from the standpoint of the chemical-rate theory," *Trans. A. I. M. E.* 143, 57 (1941).

²⁹ S. Dushman, L. W. Dunbar, and H. Huthsteiner, "Creep of metals," *J. App. Phys.* 15, 108 (1944).

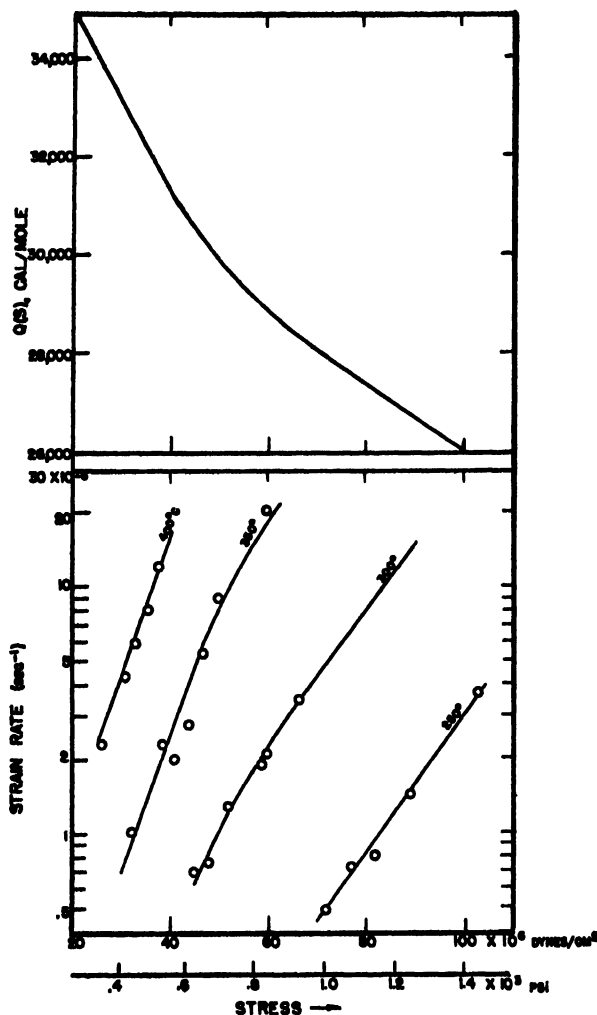


FIG. 8. Application of mechanical equation of state to pure aluminum $f_0 = 2.2 \times 10^8$. (Data after Dushman *et al.* (reference 28).)

The corresponding mechanical equation of state is therefore

$$\epsilon = f_0 \exp \left[- (Q_0 - VS) / kT \right], \quad (5)$$

where the two factors have the same interpretation as in Becker's equation.

Equations (3) and (5) are equivalent only when $S \ll S_0$, i.e., only when the stress necessary to produce deformation is much less than that necessary to produce deformation at the absolute zero temperature. Orowan has pointed out that S and S_0 are of the same order of magnitude, therefore the two equations cannot be regarded

as equivalent. A closer analysis into the details of an elementary slip process shows that both equations can only be approximations of the more general equation

$$\dot{\epsilon} = f_0 \exp \left[- Q(S) / RT \right]. \quad (6)$$

In the Becker approach the assumption is implicitly made that the probability that the thermal stress attain a certain value in a locale is independent of the stress due to external forces. In view of Orowan's concept that the external forces produce very high stresses in the locales susceptible to slip, the strict validity of this assumption is questionable. In the Condon-Kauzmann-Dushman' approach the assumption is explicitly made that the activation energy may be regarded as a linear function of the stress. Again in view of Orowan's concept, this assumption is questionable. Thus a linear correction to the heat of activation does not take into consideration that the stress changes the positions of the potential valleys and peaks as well as their heights.

In their application of Eq. (5) to experimental data, Kauzmann and Dushman found it necessary to assume that V varied with temperature, a variation for which no satisfactory interpretation was evident. As evidence of the validity of Eq. (6), Figs. 8-9 are presented, the data for which were taken from Dushman.

In their recent investigations upon the influence of changes in strain rate and in temperature on the strength properties of steels, the authors²⁰ concluded that the stress at a given strain rate depended upon strain rate and temperature only through the single parameter P defined by

$$P = (\dot{\epsilon} / f_0) e^{-Q/RT}. \quad (7)$$

It is of interest to determine under what conditions this conclusion is compatible with the general mechanical equation of state. Suppose one is interested in a range of strain rates and in a range of temperature in which the stress level is near S_0 . Upon letting ΔS represent the difference between S and S_0 , one obtains from

²⁰ C. Zener and J. H. Hollomon, "Effect of strain rate upon plastic flow of steel," J. App. Phys. 15, 22 (1944).

Eq. (6)

$$\Delta S = RT/Q'(S_0) \{ \ln(f_0/\dot{\epsilon}) - Q(S_0)/RT \}. \quad (8)$$

In the range of strain rates and temperature in which the stress level is near S_0 , the second factor contains two large terms which nearly cancel each other. A given change in temperature will hence produce a much greater relative change in the second factor than in the first factor. The effect of changes in strain rate and in temperature therefore enters primarily through the parameter P of Eq. (7) with Q taken at S_0 .

In spite of the vast amount of experimental work upon creep which has been carried out, and is still being carried out, especially in this country, no attempts have been made other than by Dushman to establish mechanical equations of state. No experimental evidence therefore exists regarding the influence upon the constant f_0 and upon the activation energy function $Q(S)$ of the various metallurgical variables, such as dissolved elements, precipitates, grain size, etc. Further, no theoretical considerations have been advanced since the early work of Orowan which gives a more detailed picture of the mechanism of plastic deformation and of the influence of thermal energy thereon.

A mechanical equation of state can exist only under conditions where the metal does not undergo significant changes with time at zero strain rate. Thus it must not soften through recovery or through recrystallization. The smaller the prior strain the higher is the temperature needed to reduce the strain hardening. The agreement of the experimental observations reproduced in Figs. 8 and 9 with the predictions of an equation of state indicate that for the metals used strain hardening was not significantly reduced by recovery or by recrystallization under the conditions of the experiment. An understanding of the conditions under which a mechanical equation of state is applicable will require the development of a theory of recovery and of recrystallization.

B. Strain Hardening

As has been previously mentioned, resistance to deformation usually increases with deforma-

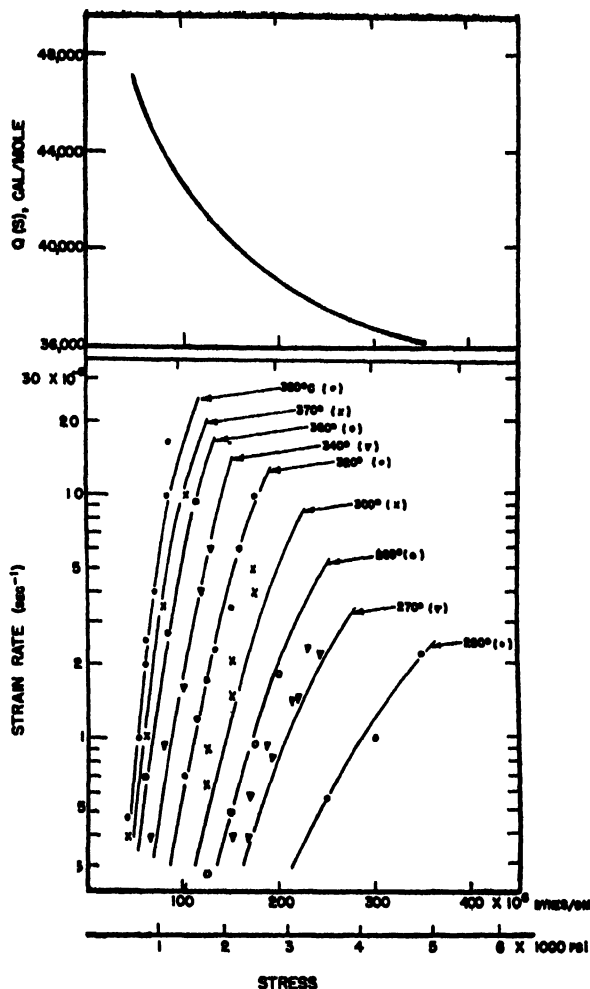


FIG. 9. Application of mechanical equation of state to an aluminum alloy with 2 percent magnesium. $f_0 = 2.5 \times 10^9$. (Data after Dushman *et al.* (reference 28).)

tion. Thus in Eq. (1), the stress S increases with the strain E . The precise manner in which the tensile stress increases with the tensile strain, at constant strain rate and temperature, has been discussed in a recent paper.³¹ It appears that, in the absence of discontinuous yielding, the stress may be represented as a power function of the strain up to strains as large as 0.4. Thus

$$S = S_0 \cdot (\epsilon/\epsilon_0)^m$$

where S_0 is the stress at some arbitrary strain ϵ_0 .

³¹ J. H. Hollomon, "Tensile deformation," *Metals Tech.* (June, 1945).

The variation of the constants S_0 and m upon the metallurgical variables has been rather exhaustively studied in the case of iron and steel. Thus in the absence of precipitates, the addition of soluble alloys³² raises S_0 without essentially altering the strain hardening exponent m . On the other hand, in steels where the strength arises primarily from insoluble carbides, the strain hardening exponent is inversely proportional to S_0 for any given carbon content. It is not known at present whether similar relations are true for other metals.

C. Generalized Stress and Strain

In the previous part of this paper it has been implicitly assumed that stress and strain referred to uniaxial tensile conditions. A discussion will now be given of the problems which arise when the stress system is of a more complex type.

Von Mises³³ has indicated the path along which the laws for uniaxial tension may be generalized to include the most general type of stress system. A certain invariant function S_I of the stresses plays the same role in determining the rate of strain as does the tensile stress in the case of uniaxial tension. This function is so defined that S_I^2 is proportional to the elastic shear strain energy. Thus

$$S_I^2 = \frac{1}{2} \{ (Y_z - Z_x)^2 + (Z_x - X_z)^2 + (X_z - Y_y)^2 \} + 3(Y_z^2 + Z_x^2 + X_z^2). \quad (9)$$

The type of strain due to a generalized stress system is of the same type as if the specimen were amorphous, thus

$$\dot{\epsilon}_{xx} = \lambda \{ X_z - (Y_z + Z_x)/2 \},$$

with analogous equations for $\dot{\epsilon}_{yy}$ and $\dot{\epsilon}_{zz}$ having the same factor λ , and

$$\dot{\epsilon}_{yz} = 3\lambda Y_z,$$

with analogous equations for $\dot{\epsilon}_{zx}$ and $\dot{\epsilon}_{xy}$. The generalized mechanical equation of state may

³² C. E. Lacy and M. Gensamer, "Tensile properties of alloyed ferrites," *Trans. Am. Soc. Metals* 32, 88 (1944).

³³ R. von Mises, "Mechanik der Festen Körper in plastischdeformablen Zustand," *Nachr. Gesellsch. Wissensch. zur Göttingen, Math-Phys. Klasse*, 1913.

therefore be written in the following manner:

$$\dot{\epsilon}_{xx} = \{ X_z - (Y_z + Z_x)/2 \} S_I^{-1} f_0 \times \exp [-Q(S_I)/RT], \quad (10)$$

with analogous equations for $\dot{\epsilon}_{yy}$ and $\dot{\epsilon}_{zz}$, and

$$\dot{\epsilon}_{yz} = 3X_z S_I^{-1} f_0 \exp [-Q(S_I)/RT],$$

with analogous equations for $\dot{\epsilon}_{zx}$ and $\dot{\epsilon}_{xy}$.

In the case of uniaxial tension it was seen that strain hardening could be taken into account by regarding f_0 and Q as functions of the strain. The same procedure may be followed in the case of a general type of stress. Before this can be done, however, it is necessary to have a measure of strain in a general type of deformation. In obtaining this measure, two steps must be taken. Firstly, the method must be determined of combining strain increments associated with increments of deformation. Secondly, strain increments must be defined in terms of increments of deformation. The first question has been answered by following a suggestion of Ludwik¹⁶ that the strain increment associated with a given increment of deformation be computed using as a reference configuration the configuration of the system before that particular increment of deformation. Thus if during an increment of deformation in uniaxial tension the length changes from l to $l+dl$, the strain increment associated therewith is dl/l , irrespective as to what prior deformation the specimen had undergone. Ludwik pointed out that the total strain is then simply the sum of all the strain increments. Thus if the gauge length changes, during a uniaxial tension, from l_1 to l_2 , the total strain associated therewith is $\ln(l_2/l_1)$. A satisfactory measure of a strain increment is an invariant function of the strain components which is proportional to S_I in an isotropic elastic medium. Such a function, $d\epsilon_I$, is defined by

$$d\epsilon_I^2 = \frac{1}{2} \{ (d\epsilon_{yy} - d\epsilon_{zz})^2 + (d\epsilon_{zz} - d\epsilon_{xx})^2 + (d\epsilon_{xx} - d\epsilon_{yy})^2 \} + \frac{3}{2} (d\epsilon_{yz}^2 + d\epsilon_{zx}^2 + d\epsilon_{xy}^2). \quad (11)$$

From the previous equations in this section it may be verified that the strain defined in this manner satisfies the following equation

$$\dot{\epsilon}_I = f_0 \exp [-Q(S_I)/RT]. \quad (12)$$

The quantities f_0 and Q are to be regarded as functions of the strain ϵ_f .

Equations (10) to (12) form a self-consistent scheme of a general mechanical equation of state. The outstanding problems relate to how well this scheme represents the behavior of actual metals, and to what deviations therefrom are to be ascribed.

One consequence of the general equation of state is that when the temperature and the strain rate are held constant a general strain hardening function exists of the type

$$S_I = S_I(\epsilon_f). \quad (13)$$

Nadai and Davis³⁴ have performed experiments designed to check this equation. Thus they measured the stress S_I as a function of strain ϵ_f under various types of stress pattern. They found that at small strains below 0.2, this equation was in fact satisfied, but that at larger strains deviations became appreciable.

In the establishment of the general mechanical equation of state in terms of the invariant functions ϵ_f and S_I , it was implicitly assumed that the material was initially, and remained, isotropic. It is to be expected that anisotropy will be introduced by plastic deformation, and that this anisotropy will be greater, the greater the deformation. The amount of anisotropy introduced by a given deformation will, however, very likely depend upon the microstructure. As an example, logarithmic strain hardening curves taken in tension and in torsion are compared for steels of two types of microstructure in Figs. 10 and 11. As might be expected, the steel which has the greatest anisotropy of structure, pearlite,

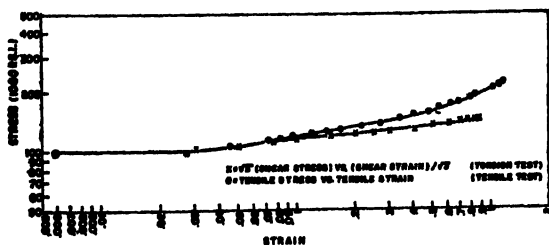


FIG. 10. Correlation of tensile and torsional tests by means of concept of generalized stress and strain (tempered martensitic steel).

³⁴ A. Nadai and E. Davis, "Plastic behavior of metals in the strain hardening range," J. App. Phys. 8, 205 (1937).

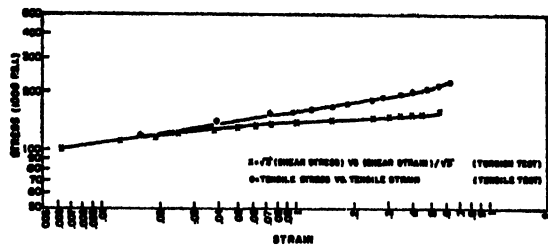


FIG. 11. Correlation of tensile and torsional tests by means of concept of generalized stress and strain (pearlitic steel).

shows the greatest apparent strain anisotropy. A further understanding of plastic deformation under general types of stress pattern must therefore involve an evaluation of the anisotropy introduced by deformation, and of the modification of the mechanical equation of state by this anisotropy.

It has been implicitly assumed that the strain and strain components in the above equation refer only to the plastic, i.e., permanent, deformation. The elastic strain components are deducible from the elastic equation, thus

$$(\dot{\epsilon}_{xx})_{\text{elastic}} = E^{-1} \{ \dot{X}_x - \sigma (\dot{Y}_y + \dot{Z}_z) \}. \quad (14)$$

The total strain components are therefore the sum of the elastic and plastic components, thus

$$(\dot{\epsilon}_{xx})_{\text{total}} = (\dot{\epsilon}_{xx})_{\text{elastic}} + \dot{\epsilon}_{xx}. \quad (15)$$

One could therefore obtain a set of equations for the time derivatives of the total strain components by using the relations analogous to Eq. (14) for the elastic part, and Eq. (10) for the plastic part. This has in fact been done by Prandtl³⁵ and Reuss.³⁶ Two factors tend to invalidate such an equation when the stress pattern is changing under conditions where the elastic strain is comparable to the plastic strain. Firstly, the anelasticity of actual metals would render ambiguous the distinction between elastic and plastic strains. Secondly, the Bauschinger effect of actual metals is very marked for plastic strain comparable to the elastic strains, but becomes negligible for much larger plastic

³⁵ L. Prandtl, Proc. 1st. Int. Cong. App. Mech. p. 43 (1924).

³⁶ E. Reuss, Zeits. f. angew. Math. u. Mech. 10, 266 (1936).

strains, as has been vividly demonstrated by Sachs and Shoji.³⁷ Prager³⁸ has applied the Prandtl-Reuss equation to a case of relaxation of residual stresses by deformation, the total strain being comparable to the elastic strains. Instead of placing the blame for lack of agreement of theoretical prediction with observation upon the anelastic and Bauschinger effects, Prager applied an empirical rule (which he claims is derived from the "Hencky-Nadai

³⁷ G. Sachs and H. Shoji, "Zug-Druckversuche an Messingkristallen," *Zeits. f. Physik* **45**, 776 (1927).

³⁸ W. Prager, "Exploring stress-strain relations of isotropic plastic solids," *J. App. Phys.* **15**, 65 (1944).

Theory") relating plastic strain and total strain which agreed with the experimental data. Analysis of his rule shows that it leads to different results when the plastic deformation takes place intermittently or continuously, contrary to all experience. In opposition to Prager's viewpoint, the authors believe an adequate understanding of the mechanics of small plastic deformations will be obtained only through a study of the basis of the anelastic and Bauschinger effects, not through attempting to find empirical formal relations between the stress pattern and the elastic and plastic strains.

Problems in Fracture of Metals*

BY J. H. HOLLLOMON AND C. ZENER

Research Division, Watertown Arsenal, Watertown, Massachusetts

In this report a review is made of the possible types of future research which might lead to an understanding of the fracture of metals, and hence to an increase of the strength level to which metals may be raised without danger of fracture. It is found that an understanding of fracture would be enhanced by research upon the following topics: 1. The effects of the following variables on the virtual fracture stress: (a) Strain; (b) strain rate; (c) temperature; (d) stress distribution; (e) structure; (f) me-

chanical history (i.e., fatigue). 2. A comparison of the fracture characteristics of metallic crystals with those of non-metallic crystals, such as rocksalt. 3. The hindrances of plastic deformation by sharp stress gradients. 4. The introduction of stress concentration by: (a) Non-metallic inclusions; (b) precipitates; (c) slip bands; (d) twin bands; (e) grain boundary deformation. 5. The effect of reversal of stresses upon the properties of slip bands.

I. INTRODUCTION

THE strength or hardness of our present-day metals cannot be further increased by known methods without inducing susceptibility to brittle failure. Therefore, practical increases in strength can only be accomplished through an understanding of the problems relating to fracture. Many features of the dynamics of crack propagation are imperfectly understood, such as the factors which determine the velocity and direction of propagation. An understanding of these features would not, however, lead to an improvement in the useful strength of metals, since a fractured metal member has failed, irrespective of the precise path or velocity of the crack. This paper will, therefore, be confined to the problems relating primarily to the initiation of fracture.

* The statements or opinions in this article are those of the authors and do not necessarily express the views of the Ordnance Department.

II. GROSS FEATURES OF FRACTURE

It has long been known that metals which are ductile under one set of conditions may break brittly under other conditions. Thus metals frequently are ductile at moderate temperatures in simple tension, but are brittle at low and at high temperatures or under a more complex stress system. It would be highly desirable to have a consistent scheme by means of which, from measurements made under one set of conditions, the fracture properties under other conditions could be predicted. Such a scheme was presented by Ludwik¹ in 1909. The cardinal feature of Ludwik's contribution was the recognition that the stress at fracture of a ductile material is the fracture stress of the material after having suffered its final strain, and not of the original material. In his scheme Ludwik con-

¹ P. Ludwik, *Elemente der Technologischen Mechanik* (Berlin, 1909).

sidered two functions of strain, as represented in Fig. 1. One function is the flow stress, defined as the tensile stress necessary to produce further plastic deformation. The other function is the virtual fracture stress, defined as the tensile stress which would be required to fracture the material if further plastic deformation did not occur. A metal deforms plastically under tension as long as the flow stress is less than the virtual fracture stress. Fracture occurs at that strain at which the flow stress curve crosses the virtual fracture stress curve. The advantage of Ludwik's scheme is that it resolves the influence upon fracture of the several variables such as stress pattern, temperature, strain rate, into the influence of these variables upon the flow stress and upon the virtual fracture stress. Its success depends upon how well the reliance of these two stress functions upon the several variables may be determined. The disadvantage of Ludwik's scheme, a disadvantage which has prevented its general adoption, is that the virtual fracture stress can be measured directly only at one strain, namely the strain at fracture. Only estimates may be obtained for strains other than this final strain.

The most straightforward method of estimating the virtual fracture stress curve was used first by Davidenkov and Wittman² and later by the authors.³ This method is applicable to all

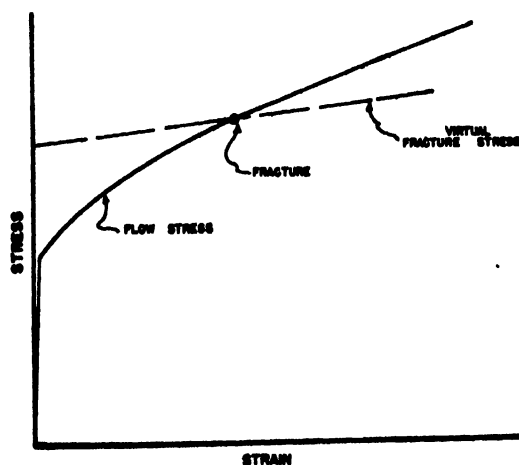


FIG. 1. Representation of fracture according to Ludwik.

² N. Davidenkov and F. Wittman, "Mechanical analysis of impact brittleness," *Tech. Phys. U.S.S.R.* 4, 3-17 (1937).

³ J. H. Hollomon and C. Zener, "Conditions of fracture in steel," *Trans. A.I.M.E.* 158, 283 (1944).

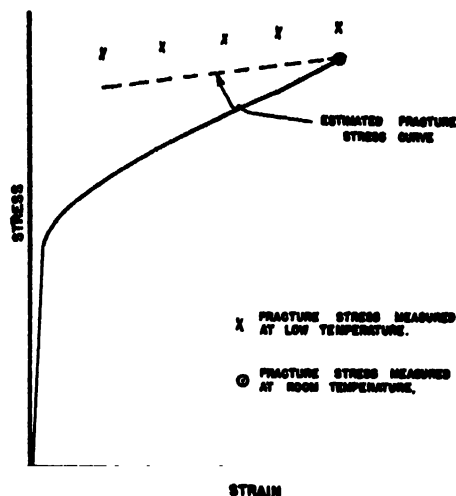


FIG. 2. Alternative method of estimating fracture stress curve.

metals which deform appreciably at room temperature but which break brittly at some low temperature. In this method specimens are deformed various amounts at room temperature, the temperature is then lowered, and they are then broken without further deformation. The fracture stress measured at the low temperature is then plotted as a function of the strain at room temperature. An estimate of the virtual fracture stress curve at room temperature is then obtained by assuming it to be parallel to the curve measured at the low temperature. The method is illustrated in Fig. 2. The method could be further improved by making the fracture measurements at a series of low temperatures, and thus at least partially justifying the undesirable assumption.

Pearlitic steels (containing lamellar carbides) and some age-hardenable alloys are brittle at low temperatures, and their virtual fracture stress curves may be obtained in this manner. Typical results for a pearlitic steel are shown as Fig. 3. All specimens deformed more than a few percent at room temperature broke with practically no further deformation at the low temperature. A specimen undeformed at room temperature, however, first exhibited an upper and lower yield point at the low temperature (Fig. 4) and actually fractured at a lower stress than it had withstood before it deformed. It can only be concluded, therefore, that at least for this type of steel* the virtual fracture stress curve has the

* Confirmatory experiments have been performed on specimens of similar steels.

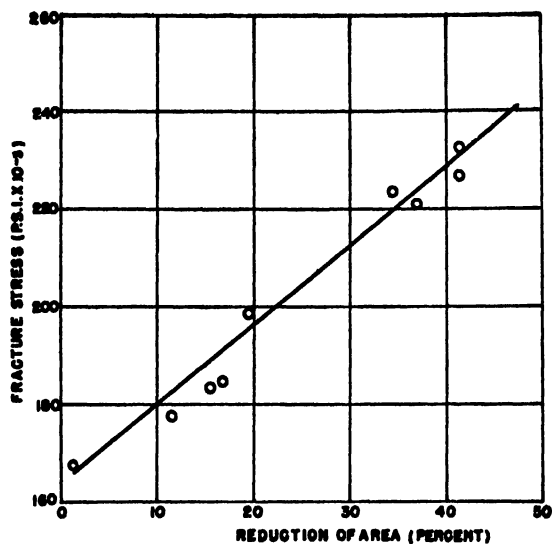


FIG. 3. Effect of prior deformation on fracture stress (at -190°C). (Pearlitic steel.)

form illustrated in Fig. 5. It is of particular interest that the estimated virtual fracture stress may be only slightly higher than the flow stress at very small strains. Any change of variables which raises the flow stress relative to the virtual fracture stress is apt to embrittle the metal.

As previously mentioned, the method described above cannot be used to estimate the virtual fracture stress curve of those metals that are not brittle at low temperatures. For such metals, one would have to utilize some other test conditions under which the metals fractured without deformation. As yet, no test conditions have been found which are as suitable in embrittling metals as is lowering the temperature (as in the case of pearlitic steels).

The virtual fracture stress curve may be highly anisotropic. The most common example of anisotropy occurs in the rolling of metal plates. The rolling of these plates raises the virtual fracture stress in the rolling direction, lowers it in the transverse direction. In an earlier paper⁴ one of the authors pointed out that longitudinal fractures of certain steels could be interpreted only in terms of an anisotropic effect of deformation on the fracture stress. The tensile strain, as already pointed out, raises the tensile stress required to fracture the specimen on a plane perpendicular to the axis. In order that a tensile

specimen can fracture longitudinally, the stress required to fracture a metal on a plane parallel to the axis must decrease sufficiently for the circumferential stress induced by necking to fracture the metal. Conversely, compression will lower the tensile stress required for fracture upon subsequent deformation in a direction opposite to the compression strain. Examples of such effects of compression are frequently found in bending. Thus, a specimen which has undergone a certain amount of bending, and which is capable of continued bending in the same direction, may break brittly when an attempt is made to unbend it. The anisotropy of the virtual fracture stress may also be visibly demonstrated by pulling specimens in tension that have previously been twisted plastically. Swift⁵ found that if the specimens had been sufficiently twisted prior to pulling, fracture occurred along a helical surface, rather than across the plane of maximum tensile stress, namely the transverse plane. Although many examples of anisotropic virtual fracture stress are known, no quantitative study has as yet been made of anisotropy. Thus, no measurements have been made to determine the relation between the virtual fracture stress and the reduction in cross section in rolling.

Under many conditions constraints prevent plastic deformation in one transverse direction.

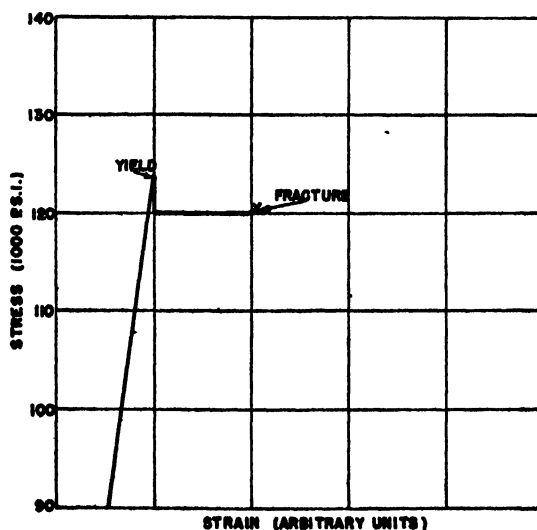


FIG. 4. Illustration of the lowering of the fracture stress by initial deformation (pearlitic steel tested at -190°C).

⁴ J. H. Hollomon, "Temper-brittleness," communicated to Am. Soc. Metals.

⁵ H. W. Swift, "Tensional effects of torsional overstrain in mild steel," J. Iron and Steel 140, 181 (1939).

Many experiments have been performed to determine the effects of a single transverse stress on the fracture stress. In order to interpret these experiments it is necessary to permit only the stress distribution to vary. Frequently in such tests other factors such as the strain to fracture varies as the stress distribution is changed. Suppose, for example, thin hollow tubes are loaded by means of a longitudinal tension and an internal pressure. Because of the presence of the transverse stress, the tensile stress required for plastic flow is raised and unless the virtual fracture stress curve is identically affected the strain to fracture will also change as illustrated in Fig. 6. Furthermore, as the circumferential tension in the hollow tubes becomes larger than the longitudinal tension, the fracture may change from transverse to longitudinal. Because of the anisotropy, the fracture stress may be altogether different in the new direction, and more important, it may depend differently upon strain. Experiments of Siebel and Maier⁶ on tensile tests of hollow cylinders under internal pressure indicate that a single transverse stress has no effect upon the virtual fracture stress, provided the assumption is made that in their single phase materials the virtual fracture stress was independent of strain. Indirect evidence also exists that the transverse stress associated with such restraints does not influence the virtual fracture stress. Thus, the brittle failure of pearlitic steels in the notched-bar impact test can be interpreted

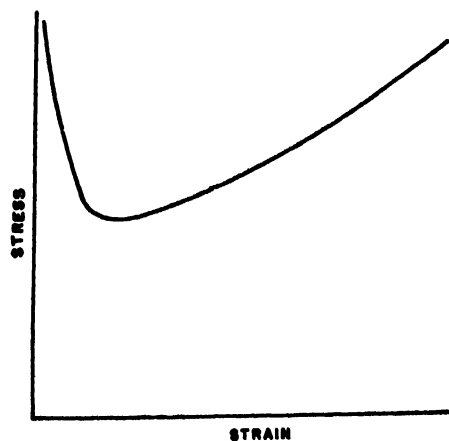


FIG. 5. Proposed shape of virtual fracture stress curve (for pearlitic steels).

⁶ E. Siebel and A. Maier, "Effect of multiaxial stress states on the ductility," *Zeits. f. V. D. Ing.* 77, 1345-1349 (1933).

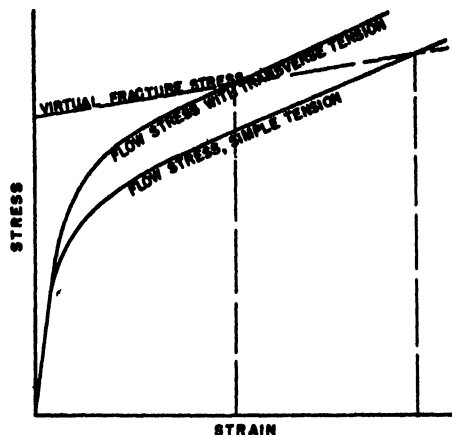


FIG. 6. Possible effect of transverse stress on flow curve and fracture strain (if virtual fracture stress is unaffected).

only by assuming that a single transverse stress has much less effect upon the virtual fracture stress than upon the flow stress.⁷ It would be highly desirable to have direct evidence upon the variation of the virtual fracture stress with a single transverse stress.

Under less frequent conditions a biaxial transverse tension is present. By testing tensile specimens having circumferential notches of various depths and radii, Kuntze,⁸ McAdam,⁹ and Sachs¹⁰ have attempted to evaluate the influence of a biaxial transverse tensile stress upon the virtual fracture stress, called by them the "technical cohesion strength." In their measurements the average longitudinal stress at fracture is taken as the fracture stress. The interpretation of the variation of this stress with variations in testing conditions is usually extremely difficult. As discussed by these workers, the effects of the three separate variables, strain, stress complex, and variation of stresses throughout the specimen, may be reflected by these measurements. In one case Sachs and his collaborators were able

⁷ J. H. Hollomon, "Notched-bar impact test," *Trans. Am. Soc. Min. Met. Eng.* 158, 298-327 (1944).

⁸ W. Kuntze, for survey and bibliography, see D. J. McAdam, *Trans. Am. Inst. Min. Met. Eng.* 150, 311 (1942).

⁹ D. J. McAdam, "Technical cohesive strength of metals in terms of the principal stresses," *Metals Tech.* (December, 1944). References given to prior work.

¹⁰ G. Sachs, J. D. Lubahn, and E. J. Ebert, "Effects of notches of varying depth on the strength of heat treated low alloy steels," *Trans. Am. Soc. Metals*, 34, 517-544 (1945); G. Sachs and J. D. Lubahn, "Effect of triaxiality on the technical cohesive strength of steels," A. S. M. E. preprint, 1945.

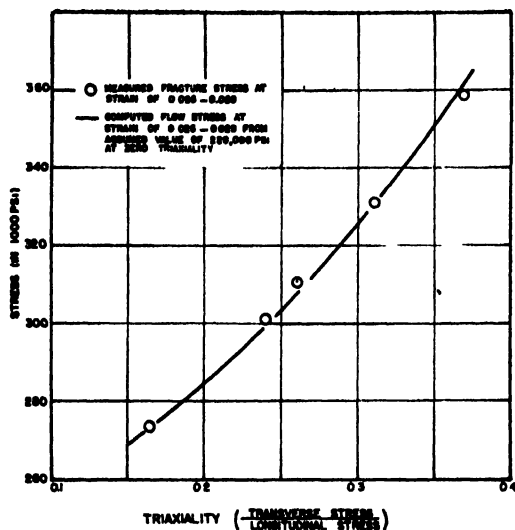


FIG. 7. Dependence of fracture stress upon biaxial tension (after data of Sachs (reference 10)).

to obtain the same strain (0.026 to 0.028) with several notch depths and notch radii. In this case the strain variable was absent. If it is assumed that the stress distribution is essentially constant throughout the unnotched region with notches of various depths and radii, then the observed change in fracture stress may be regarded as a reflection of the triaxiality of stress (ratio of transverse to longitudinal stress). Their results are reproduced as Fig. 7. The self-consistency of the data may be seen from the close agreement of the measured fracture stress with the computed flow stress at fracture. The rise of the virtual fracture stress with biaxial tension, as demonstrated in Fig. 7, confirms the earlier conclusions of McAdam and his collaborators. The constancy of the strain to fracture implies that in this particular case the virtual fracture stress rises just as rapidly with biaxial tension as does the flow stress, and hence that a biaxial tension has no embrittling effect. This conclusion is in marked disagreement with current ideas upon the influence of triaxial stresses. Investigation of the type given by Sachs should therefore be further pursued.

Fracture in steels is most apt to occur under impact conditions at low temperatures. Both a lowering of temperature and an increase in the rate of strain are known to raise the flow stress of all steels. Experiments⁹ have indicated that the virtual fracture stress of all steels likewise

rises with a lowering of the temperature, but not quite as rapidly as does the flow stress. It is the resulting relative rise of the flow stress with respect to the virtual fracture stress that leads to brittle failure in the case of pearlitic steels. Because of the relatively small effect of changes of strain rate obtainable with the usual type of testing machines, combined with the relatively large scatter in fracture measurements, no conclusions can as yet be drawn upon the influence of strain rate upon the virtual fracture stress.

Nearly brittle fracture of metals is also likely to occur at high temperatures under certain conditions. As was first observed by Rosenhain and Archbutt,¹¹ these conditions are such as to favor grain boundary slip. Under these conditions the virtual fracture stress may be considered as decreasing with time, although the accompanying strain may be negligible. This concept of the virtual fracture stress decreasing with time suggests interesting experiments which have as yet not been performed. Thus, if after the virtual fracture stress of a specimen is lowered by the application of a stress at an elevated temperature, the specimen is lowered to room temperature without removing the stress, the virtual fracture stress should stay lowered, and may thus result in brittle fracture at room temperature.

III. MECHANISM OF FRACTURE

It has long been recognized that the observed fracture stress of metals, in fact of all crystalline substances, is at least two orders of magnitude lower than the intrinsic fracture stress as computed on the basis of the forces between atoms. The only way to reconcile this divergence between observed and theoretical fracture stress is through the assumption that at any one instant the material fractures only in localized regions, regions in which the stress is much higher than the average applied stress.

Stress concentrations in localized regions were first investigated in detail by Inglis.¹² He pointed out that the tensile stress at the base of a notch

¹¹ W. Rosenhain and S. L. Archbutt, "On the inter-crystalline fracture of metals under prolonged application of stress," *Proc. Roy. Soc.*, 95, 55 (1919-20).

¹² C. E. Inglis, "Stress in a plate due to the presence of cracks and sharp corners," *Trans. Inst. Naval Architects* 55, Pt. 1, 219 (1913).

may be much higher than the average tensile stress throughout the specimen, the stress concentration being given by $1+2(a/\rho)^{1/2}$, where a is the depth of the notch, ρ the radius of curvature at the base of the notch. Joffé¹³ and his collaborators have found that the discrepancy between the usually observed fracture stress of rock salt crystals and the theoretical fracture stress may be accounted for entirely in terms of surface notches in the form of small surface cracks. By immersing these specimens in water and thereby continually dissolving away the surface, they were able to raise the actual fracture stress by a factor of 400. Under these conditions the fracture stress is in fact very nearly equal to the theoretical value.

Well-prepared metal specimens appear to be free of surface defects of the type found in dry rock salt crystals. Their weakness with respect to fracture must be blamed upon defects within the material itself, for once deformation (in tension) has proceeded to such an extent that necking has commenced, fracture begins in the interior of the specimen and propagates outwards. That fracture should commence in the interior of the necked specimens is understandable in view of the analysis by Bridgman¹⁴ of the stress system within a necked region. In this stress system the tensile stress is a maximum along the axis, a minimum at the surface. It would be expected that metal crystals of the same degree of purity as Joffé's rock salt crystals would manifest the same relatively high fracture stress. As yet no experimental work has been done to check the correctness of this viewpoint.

The relatively low fracture stress of actual metals may most readily be interpreted in terms of the presence of micro-cracks within the metal itself. Starting from the observation that a crack will propagate only if the total free energy of the system is lowered by the propagation, Griffith¹⁵ deduced that a circular crack of a given radius a will propagate only when the tensile stress normal to the crack exceeds a

certain critical value S . Griffith's formula is:

$$S = (\sigma G/ca)^{1/2}. \quad (1)$$

Here σ is the surface energy per unit area, G is the shear modulus, and c a constant of the order of magnitude of unity. Griffith developed his theory, and obtained experimental confirmation of it only for amorphous substances. The resistance to plastic deformation of such substances increases rapidly as the temperature is lowered, so that at sufficiently low temperatures, room temperature for ordinary glass, a crack may propagate unaccompanied by any plastic deformation. In amorphous substances, therefore, where the new surface produced by crack propagation has not undergone plastic deformation, the appropriate value of surface energy σ in Eq. (1) is very nearly the surface energy of the substance in the molten state. This surface energy may be readily measured. The situation is more complex in the case of crystalline substances. On one hand the resistance to plastic deformation of crystalline substances increases relatively slowly with temperature, so that such deformation cannot usually be eliminated entirely merely by lowering the temperature. On the other hand, the laws of plastic deformation obtained by macroscopic measurements are not applicable when the stress varies appreciably in a distance less than the linear dimensions of a single crystal. A stress greatly in excess of the macroscopic yield stress will not necessarily lead to plastic deformation if the stress is confined to a sufficiently localized region. No investigation has as yet been made to determine under what conditions a sharp crack will propagate without any attendant plastic deformation. If plastic deformation does occur, the energy associated with it must be added to the surface energy σ in order that Eq. (1) be applicable.

Although the ambiguity which now exists regarding the proper value of the surface energy σ prevents the quantitative application of Eq. (1) to metals, the qualitative aspects of Griffith's ideas may be used to correlate many diverse fracture phenomena.¹⁶ These phenomena are associated with the reorientation and distortion of the micro-cracks by deformation. Figure 8

¹³ A. Joffé, "Plastizität und Festigkeit der Kristalle," Proc. 1st Int. Cong. Ap. Mech., Delft (1924.)

¹⁴ P. W. Bridgman, "Stress distribution at the neck of a tensile specimen," Trans. Am. Soc. Metals 32, 553 (1944).

¹⁵ A. A. Griffith, "Phenomena of flow and rupture in solids," Phil. Trans. Roy. Soc. 221, 163 (1920).

¹⁶ C. Zener and J. H. Hollomon, "Plastic flow and rupture of metals," Trans. Am. Soc. Metals 33, 163 (1944).

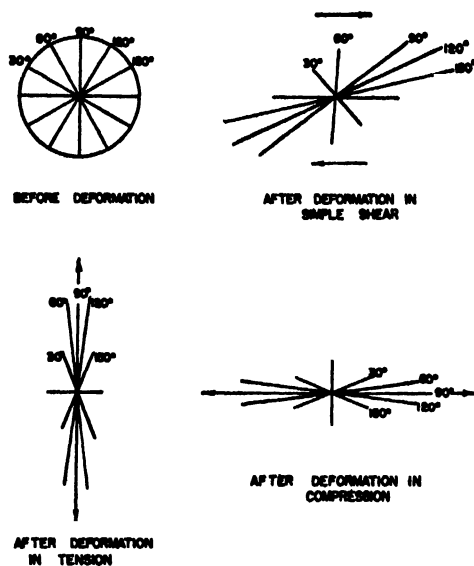


FIG. 8. Reorientation of micro-cracks by deformation.

illustrates how an initially random distribution of micro-cracks acquires a highly anisotropic distribution by deformation. Thus in tension the cracks are elongated along the tensile axis, are shortened in the transverse direction. The critical tensile stress necessary to cause some of the micro-cracks to propagate, and thus lead to rupture, is raised by deformation when the stress is longitudinal, lowered by deformation when the stress is transverse. The most common example occurs in the rolling of plate, where the fracture stress is raised in the rolling direction but reduced in the transverse direction. Another example occurs in the usual tensile tests of certain steels in which the tensile strain is found to raise the fracture stress perpendicular to the axis of the specimen but to lower it on a plane parallel to the axis. In fact, as mentioned earlier, the fracture stress with respect to transverse tension may become so low that the transverse tensile stress within the neck of a tension specimen¹⁴ causes the specimen to rupture longitudinally rather than transversely. The most spectacular example of the influence of micro-crack reorientation upon fracture occurs when a specimen is first plastically twisted and then pulled in tension. The spiral-type crack which is observed is just of the type which one would predict from an examination of reorientation effects.¹⁵

Although the reorientation effects of the micro-cracks are readily understood, considerable

uncertainty exists regarding the nature and origin of the micro-cracks. Inclusion of non-metallics, such as iron oxides and sulphides in steel, no doubt gives rise to stress concentration, and might, for certain purposes, be considered as the micro-cracks responsible for fracture. It is thus known that the "dirtier" the steel, i.e., the more non-metallics it contains, the worse are its fracture properties. It is also known that hot working introduces reorientation effects that are not removed by heat-treatment involving recrystallization of the matrix; reorientation effects are thus interpretable only in terms of the non-metallic inclusions. The fracture properties cannot, however, be described entirely in terms of the non-metallic distribution, for it is known that the structure of the matrix also has a marked influence upon the fracture stress. Thus, in steel, the tempered martensitic structure (spheroidal carbides) has a higher fracture stress at small strains³ than a pearlitic structure (lamellar carbides) having the same resistance to plastic deformation; for a given type of structure the fracture stress is higher the smaller the carbide particles. The matrix itself must, therefore, contain origins for stress concentration. The stress concentration at any point within the matrix is then the product of two stress concentration factors, that arising from the dirt and that arising from the matrix itself. As yet no experimental study has been undertaken from which the relative magnitude of these two factors may be ascertained.

Considerable uncertainty exists as to the origin of the stress concentrations within the matrix itself. If the matrix contains precipitated particles, such as carbides in steel, these particles will be a source of some stress concentration. The stress concentration factor will depend upon the shape and size of the particles. Thus, as previously mentioned, plate-like particles give rise to higher stress concentrations and, therefore, lower fracture stresses than spheroidal particles. Evidence will shortly be presented by one of the authors (JHH) that the larger the size of the carbide particles, the higher the stress concentration factor associated therewith. The data are not sufficiently complete, however, that precise relations can be derived.

Quite apart from the presence of precipitated

particles, the matrix acquires stress concentrations through the very act of plastic deformation. As first pointed out by Orowan,¹⁷ stress concentrations arise at the edges of slip bands. One of the authors¹⁸ has cited examples of fracture that can be interpreted only through the assumption that such stress concentrations may increase with time under a constant applied stress, just as if the slip bands behaved in a viscous manner. It might be expected that such stress concentration would be larger the larger the lateral extent of the slip bands, and therefore, the larger the grain size. As yet this influence of grain size *per se* has not been demonstrated. It is generally recognized that steels of small grain size have better fracture properties than steels of large grain size. In steels, however, the grain size has a marked effect upon the manner in which the carbides precipitate, and so the direct effect of grain size cannot be determined. The stress concentration associated with plastic deformation has the unique property that it is not present until after some plastic deformation has occurred. It is thus expected that the fracture stress of well-annealed specimens would be higher before than after a slight plastic deformation. The data presented above as Figs. 4 and 7 may be regarded as further confirmation of the viewpoint that the virtual fracture stress of the undeformed material may be very high. Further corroboration that fracture stress is intimately related to plastic deformation may be found in the observation that the fracture stress varies with strain rate and with temperature in approximately the same manner as does the resistance to deformation. No experiments have so far been undertaken to determine just how large a stress an undeformed specimen would sustain without fracturing.

That stress concentrations are built up through plastic deformation itself gives rise to the possibility that the fracture properties may be modified through alterations in the conditions of test. An example has recently been described by Bridgman.¹⁹ He found that when two identical specimens are deformed in tension by the same amount, one under atmospheric pressure and one under high pressure, the latter will suffer

much more additional deformation under atmospheric pressure than the former specimen, and hence has a considerably higher fracture stress. It appears as if the high pressure prevents high stress concentrations from opening micro-cracks. Initial deformation at an elevated temperature might have the same effect provided the temperature is not so high that recovery or recrystallization occurs. At the higher temperature a smaller stress would be required to produce the same extension as at room temperature, and hence fewer micro-cracks might be formed. Experiments to demonstrate this effect of the temperature of prior deformation upon fracture properties have not yet been performed. Possibly the formation of twin bands can give rise to greater stress concentrations at their edges than the formation of slip bands, for some experiments indicate that twin bands may be associated with brittle failure.²⁰

As early as 1920 Rosenhain¹¹ and his collaborators presented convincing evidence that grain boundaries behave in a viscous manner, and therefore that low rates of deformation and elevated temperature favor slipping at the boundaries, and that this slipping leads to premature fracture. Although not explicitly stated by Rosenhain, it is evident that the premature fracture arises from stress concentrations caused by the grain boundary slip. Unfortunately the concept of the viscous behavior of grain boundaries has not been extensively utilized by students of metals in this country. As a consequence no advance has been made during the intervening twenty-five years towards an evaluation of those factors that influence the stress concentration arising from the viscous slip at grain boundaries. Many interesting experiments which have not as yet been carried out are suggested by this concept. For example, the stress concentration induced by applying a stress at a high temperature can be frozen in by cooling to room temperature before removal of stress, and the metal thereby embrittled.

While a specimen may be able to sustain a certain stress for an indefinite time, it may fracture if the stress is alternately reversed and reapplied a large number of times. This fracture

¹⁷ E. Orowan, "Zur Kristallplastizität," *Zeits. f. Physik* 89, 605, 614, 634 (1934).

¹⁸ C. Zener, "Anelasticity of Metals" (in print, A.I.M.E.).

¹⁹ P. W. Bridgman, *Rev. Mod. Phys.* 17, 3 (1945).

²⁰ N. Davidenkov, *Dynamic Tests of Metals* (in Russian), 1936.

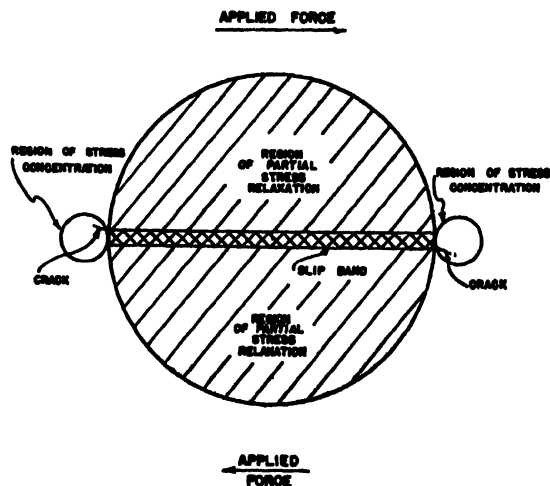


FIG. 9. Representation of stress pattern in matrix adjacent to a freshly formed slip band.

by an alternating stress, called fatigue, has been known for over 80 years. The practical importance of fatigue has led to the accumulation of a multitude of data for various metals relating the number of cycles necessary for fracture to the peak applied stress under various conditions. In this country little work has been undertaken for the purpose of finding the mechanism of fatigue failure. At the National Physical Laboratory in England, an extensive descriptive study has been made of the initiation and propagation of fatigue cracks in single crystals and in polycrystals. The early work, including the progressive changes in the hysteresis loops, has been reviewed by Moore and Kommers.²¹ The descriptive work in the National Physical Laboratory initiated by Ewing²² and collaborators, and continued by Gough and collaborators, upon the initiation of fatigue cracks, and upon their relation to slip bands, has been reviewed by Gough.²³ No attempt has been made to interpret the observed results in terms of the micro-structure, other than through the concept of attrition introduced by Ewing and Humfrey,²² according to which the slip bands become more and more disorganized as the cyclic stressing proceeds.

A cursory analysis of the stress pattern in the

²¹ H. F. Moore and J. B. Kommers, *The Fatigue of Metals* (McGraw-Hill Book Company, Inc., New York, 1927).

²² J. A. Ewing and J. C. W. Humfrey, "Fracture of metals under repeated alterations of stress," *Phil. Trans. Roy. Soc.* 200, A241 (1903).

²³ H. J. Gough, "Crystalline structure in relation to failure of metals—especially by fatigue," *Proc. A. S. T. M.* 36, Pt. 2, 3 (1933).

neighborhood of a slip band leads to a qualitative interpretation of the salient features of fatigue. Such a stress pattern is illustrated in Fig. 9 for the case after a slip band has formed but before the applied stress is removed. The shear stress in the regions on either side of the slip bands has become partially relaxed. On the other hand, the shear stress has become highly concentrated in the region immediately adjacent to the ends of the slip bands. The high tensile stresses associated therewith lead to the formation of a small crack, as indicated in Fig. 8. This crack will not propagate outside of the zone of high stress concentration, for its size is too small to be propagated by the applied stress. Since the dimensions of the zone of the high stress concentration are comparable to the width of the slip band, the length of the crack will likewise be comparable to the width of the slip band. When the applied stress is now reversed, those regions in which the stress had formerly relaxed will now have a higher stress than the applied stress. If the original slip band is not susceptible of further deformation, the adjacent region will now slip in the opposite direction, again partially relieving shear stress in adjacent regions. Each successive alteration of stress thus increases the width of the slip band, and hence also the length of the cracks at edges of slip bands. When the cracks have attained sufficient length, the specimen fractures.

The above description of a slip band under fatigue conditions shows that the slip band must necessarily broaden. This has in fact been observed.¹⁸ The above description also indicates that the progress of fatigue is intimately associated with residual stresses in the region surrounding the slip bands. If these stresses could be relieved by some process, the contribution to fatigue of all prior slipping could be eliminated. It appears that the harmful effects of over-stressing, and their elimination by under-stressing, are explicable in terms of these residual stresses. The above description in fact suggests a multitude of experiments that have not as yet been undertaken. Examples are: the measurement of fracture stress before fatigue failure as a function of number of stress reversals; elimination of harmful effects of prior fatigue by application of a set of reversals of slowly decreasing amplitudes.

Some Integral Equations of Potential Theory

BY H. BATEMAN

California Institute of Technology, Pasadena, California

1. INTRODUCTION

INTEGRAL equations arise either directly or methodically. The direct formulation of a problem by means of assumed or recognized principles frequently involves integration and may lead to an integral equation in which the unknown function has a simple physical meaning. When, for instance, use is made of the superposition of effects, one of the factors in the elementary law may be known while the other is unknown and is to be found from the equation set up. For example, one factor may be a force and the other a density.

When an integral equation arises methodically in the analysis of a problem, the unknown function or functions may be auxiliary and needed only for the ultimate expression of the physical quantities. When it happens that the auxiliary functions do have a direct physical meaning, the method gives the principles for a direct treatment. One fruitful method is to introduce all the supplementary conditions when definite integrals involving arbitrary functions are used to represent suitable types of solution of the partial differential equations of a problem.¹

2. INVERSE PROBLEMS

In the theory of attractions either the law of force or the nature of the attracting matter may be sought from observations or field data. Progress has generally been obtained by making assumptions and analyzing their consequences in an endeavor to account for the facts. A simple consequence of the Newtonian law of attraction between particles was used, for instance, by Joseph Priestley in a deduction of the law of force between electrified particles from a known property of an electrified conductor. This is pointed out by Whittaker.² The effect of a conductor in modifying the action of a charged body

also provided an inverse problem in which the aim was to find the induced surface distribution of electricity on the conductor which was regarded as responsible for the effect. This problem was ultimately formulated as an integral equation.

Problems in the theory of attractions relating to the figure of the earth were ultimately formulated as integral equations. For the history and terminology of these problems reference may be made to Bulletin 78 of the National Research Council and particularly to the article of W. D. Lambert.³ When it is desired to derive the form of the geoid from the gravitational field of the earth, an integro-differential equation is derived whose solution is not unique as was noted by Brillouin⁴ and other writers. Mineo⁵ made the solution unique by adding certain assumptions. When these are of the type used by Stokes, the integral equation is of Fredholm's type and had been noted previously by Signorini.⁶ Tables for determining the form of the geoid and its indirect effect on gravity are provided by Lambert in the chapter mentioned and in a later work in which he collaborated with F. W. Darling.⁷ A Fredholm equation was found also by Idelson⁸ who used the potential of a simple layer (the disturbing layer) in an equation of Helmert. For the equation of Fredholm the well-known methods of Neumann, Fredholm, Hilbert, and Schmidt are available. One good feature of these methods is that the inversion formulae generally involve integrations rather than differentiations which are hard to make accurately when data are few

¹ See W. Lambert, *The Shape and Size of the Earth* (Nat. Acad. Sci., Washington, D. C.), Chapter IX.

² M. Brillouin, *Comptes rendus* 180, 987-992 (1925); 181, 749-752 (1925).

³ C. Mineo, *Palermo rendus* 51, 293-303 (1927).

⁴ A. Signorini, *Acc. dei Lincei* [5] 20, 154-160, 219-222 (1911). See also N. Moiseiev, *Gerlands Beitr. z. Geophys.* 42, 279-290 (1934); B. Saltykov, *Acad. Sci. USSR, Comptes rendus (Doklady)* [2] 16, 133-140 (1937); *Jahrb. Fortschritte Math.* 63, 407 (1937).

⁵ W. D. Lambert and F. W. Darling, *U. S. Coast Geod. Surv., Spec. Pub.* 199, 137 (1936).

⁶ N. Idelson, *Gerlands Beitr. z. Geophys.* 40, 24-28 (1933).

¹ H. Bateman, Report, British Assoc. Adv. Sci., Sheffield, 4060407 (1910).

² E. T. Whittaker, *History of the Theories of Aether and Electricity* (University Press, Dublin, 1910), p. 50.

and inexact, yet the numerical work with the inversion formulae is often very great.

3. DATA ON A VERTICAL LINE

Joachimstahl⁹ in 1861 sought a law of force such that the potential of a uniform distribution of matter along a line may be an assigned function of the distance of an attracted particle from the line. His equation resembled Abel's integral equation and was solved by a known inversion formula. Beltrami¹⁰ also made a clever use of the inversion formula for Abel's equation to solve a mixed problem of potential theory in which a symmetrical potential has given values over a circular disk and the normal derivative is zero over the part of the disk's plane outside the disk. This is really a particular case of the problem of Dirichlet—to find a potential function, regular in a region, from assigned values on the boundary of the region—the surface values assigned on the two faces of the disk being such that the desired potential function is symmetrical about the plane of the disk.

Dhar¹¹ discussed Joachimstahl's problem and formulated a second problem with vertical line data in which the aim was to find a rectilinear layer symmetrical about a point O on the carrying line and such that, with the Newtonian law of force, the attraction at a point P is an assigned function $f(h)$ of the distance $OP(=h)$, when OP is normal to the line. The integral equation for

⁹ F. Joachimstahl, *J. reine angew. Math.* **58**, 135–137 (1861).

¹⁰ E. Beltrami, *Acad. Sci. Bologna, Mem.* [4] **4**, 211–246 (1882), *Opere*, v. 4, 45–67. An account of Beltrami's work is given in A. G. Webster, *Partial Differential Equations of Mathematical Physics*, edited by S. J. Plimpton (G. E. Stechert and Company, New York, 1933), second edition, pp. 368–375. For a good account of the theory of Abel's integral equation see M. Bôcher, *An Introduction to the Study of Integral Equations* (Cambridge University Press, New York, 1909). In the case of Joachimstahl's equation

$$f(h) = \int_h^\infty \frac{F(r)dr}{(r^2 - h^2)^{1/2}};$$

the inversion formula,

$$F(r) = -(2/\pi) \int_h^\infty \frac{rf'(h)dh}{(h^2 - r^2)^{1/2}},$$

is derivable from the results mentioned by Bôcher by a simple substitution. A good survey of the literature on equations resembling Abel's equation is given by H. T. Davis, "A survey of methods for the inversion of integrals of Volterra type," *Indiana University Studies* **14**, Nos. 76, 77 (1927).

¹¹ S. C. Dhar, *Bull. Calcutta Math. Soc.* **10**, 151–156 (1918–1919).

the linear density $F(t)$ is

$$f(h) = 2h \int_0^\infty (h^2 + t^2)^{-1/2} F(t) dt, \quad (3.1)$$

and is of a well-known type because $\frac{1}{2}f(h)$ can be regarded as the value on the line of symmetry OP of a symmetrical potential V whose value on the plane ($z=0$) normal to OP is $F(r)/r$, where r is the distance from O . The integral is then a special case of the integral of Poisson for the solution of the Dirichlet problem for a region bounded by a plane. The inversion formula

$$F(t) = (t/2\pi) \int_0^\pi f(it \cos a) da \quad (3.2)$$

is sometimes useful but requires a continuation of $f(h)$ to complex values of h and, perhaps, a derivation from the data of the derivatives of the function f . A similar remark applies when use is made of power series in ascending or descending powers of h and the associated series of Legendre functions. The data must be plentiful and accurate to permit a determination of the derivatives or series.

The use of vertical line data has been developed by some members of the Gulf Research and Development Company and a partial account of some of the methods that are used by this and other companies is given in papers of Evjen¹² and Hughes.¹³

The inversion formula

$$f(h) = 2 \int_0^\infty e^{-hz} dz \int_0^\infty J_0(zt) F(t) dt \quad (3.3)$$

makes the determination of $F(t)$ depend on the solution of integral equations of well-known types, $J_0(w)$ being the Bessel function of zero order. A practical method of solving the equation with exponential integral is developed in this paper.

4. IMPROVEMENTS IN OBSERVATIONAL METHODS

A description of some of the instruments used in gravity work is given in Bulletin 78 already mentioned, but the invention of instruments of

¹² H. M. Evjen, *Geophys.* **3**, 72–95 (1938).

¹³ D. S. Hughes, *Geophys.* **7**, 169–178 (1942).

precision during the last 25 years has placed much more valuable equipment within the reach of men interested in observations of gravity. The invention by Baron Roland von Eötvös¹⁴ of a torsion balance and of a more general apparatus permitting a determination of the curvature of the lines of force of the weight field of the earth did much to arouse interest in accurate observations and in gravity methods of geophysical prospecting for, as is remarked for instance by Roever,¹⁵ local irregularities, such as minerals and mountains, may produce curvatures much larger than those given by the potential function whose level surface is the standard spheroid of geodesy. A brief history of the development of the gravity method of prospecting for oil is given by Eckhardt.¹⁶ It seems that this work started in Europe about 1915 and in the United States about 1924. The subject is discussed also by Shaw and Lancaster Jones.¹⁷ An instrument like the Eötvös balance but much smaller was used by Jones and Davies¹⁸ for gravity measurements. It was thought that a handy instrument would facilitate observations giving valuable information concerning the shape and dimensions of buried anticlines. Many other instruments embodying pendulums or springs have been devised for measurements of gravity. An account of some of these has been given by Hoskinson.¹⁹

Eckhardt says that in the period 1930-1935 the torsion balance began to be superseded by the gravimeter. For the evolution of this instrument reference may be made to the papers of Skeeters,²⁰ Mott-Smith,²¹ Hammer,²² Neuman,²³

Bruckshaw,²⁴ Welch,²⁵ Clewell,²⁶ Fenwick,²⁷ McCollum and Brown,²⁸ Heiland,²⁹ and many others. In 1937 Stubbe and Schmidt³⁰ compared the gravimeter with the torsion balance and concluded that the latter was preferable, but Thyssen,³¹ who can speak with some authority, disagrees. For other types of balance or observing instrument references may be made to papers of Duffield,³² Soler,³³ Vening Meinesz,³⁴ Lancaster Jones,³⁵ Tomaschek and Schaffernicht,³⁶ Holweck,³⁷ and numerous papers by A. Schleusner, S. v. Thyssen, and others.³⁸

5. HORIZONTAL LINE DATA

Although some of the inverse problems of potential theory had been stated and partially solved before 1930 when many workers became interested in geophysical prospecting, the actual use of the methods of integral equations was delayed because the data were not sufficiently accurate. With the advent of instruments of precision the situation was somewhat changed, and some members of the staff of the Gulf Research and Development Company began to study a very simple integral equation in order to acquire some skill in the solution of inverse potential problems. A solution of this guiding integral equation in the form of a double integral of Fourier type was obtained by H. M. Evjen but was not regarded as suitable for purposes of calculation. Evjen then wrote to the present author to ask if his method was new and if another method of solution was available which would involve, say, only one integration.

¹⁴ J. M. Bruckshaw, *Proc. Phys. Soc. London* **53**, 449-465, discussion 465-467 (1941).

¹⁵ G. I. Welch, *Rev. Sci. Inst.* **12**, 179-181 (1941).

¹⁶ D. H. Clewell, *Geophys.* **7**, 155-168 (1942).

¹⁷ W. H. Fenwick, *Mines Mag.* **32**, 504-507, 509 (1943).

¹⁸ E. V. McCollum and A. Brown, *Geophys.* **8**, 379-390 (1943).

¹⁹ C. A. Heiland, *Geophys.* **8**, 119-133 (1943).

²⁰ G. Stubbe and K. H. Schmidt, *Petroleum Eng.* **8**, 137-139 (1937).

²¹ S. v. Thyssen, *Petroleum* **23**, 34-36 (1937).

²² W. G. Duffield, *Roy. Astron. Soc. Geophys. Suppl.* **1**, 161-204 (1924).

²³ E. Soler, *Washington Acad. Sci.* **16**, 261-266 (1926).

²⁴ F. A. Vening Meinesz, *Theory and Practice of Gravity Observations at Sea* (Waltman, Delft, 1929).

²⁵ E. Lancaster Jones, *J. Sci. Inst.* **9**, 341-353, 373-380 (1932).

²⁶ R. Tomaschek and W. Schaffernicht, *Ann. d. Physik* **5**, 15, 787-824 (1932).

²⁷ F. Holweck, *Bull. Geodésique* **46**, 295-303 (1935).

²⁸ A. Schleusner, S. v. Thyssen, *et al. Beitr. z. angew. Geophys.* **4-6**.

¹⁴ R. von Eötvös, *Konferenz der Erdmessung, Budapest* **1**, 337-395 (1906). *Ann. d. Physik* **68**, 11-66 (1922). See also W. Haubold, *Zeits. f. Geophysik* **8**, 446-453 (1932).

¹⁵ W. H. Roever, "The Weight Field of the Earth," *Washington University Studies, new series, Science and Technology*, No. 11 (1940).

¹⁶ E. A. Eckhardt, *Geophys.* **5**, 231-242 (1940).

¹⁷ H. Shaw and E. Lancaster Jones, *Proc. Phys. Soc. London* **35**, 151-166, 204-212 (1923).

¹⁸ J. H. Jones and R. Davies, *Roy. Astronom. Soc. Geophys. Suppl.* **2**, 1-32 (1928).

¹⁹ A. J. Hoskinson, *Trans. Am. Geophys. Union*, 44-45 (1936).

²⁰ W. W. Skeeters, *Mines Mag.* **29**, 304-309 (1939); **31**, 127-129, 472, 495 (1941).

²¹ L. M. Mott-Smith and F. W. Mott-Smith, *Petroleum Eng.* **10**, 85-86, 89-90, 92, 94, 96-97 (1939); L. M. Mott-Smith, *World Petroleum* **11**, 64-67 (1940).

²² S. Hammer, *Geophys.* **4**, 184-194 (1939).

²³ G. Neuman, *Zeits. Bohrtechniker* **58**, 33-37, 53-58 (1940).

As a matter of fact a solution of the desired type had been found by Chester Snow³⁹ in 1923. He had obtained the integral equation (with a restriction on the unknown function, which was, in his case, odd) in a spectroradiometric analysis of radio signals. Snow derived the solution also in the form of a double integral but reduced this to a single integral and extra terms by an operational method. The result was found also by means of Cauchy's theorem and some remarks were made on numerical methods.

In a reply to Evjen's letter the present author gave a method of analytical continuation which led naturally to the simple form of solution. This method seems to be of wide application in inverse potential problems in which the assigned function is analytic in the sense of Weierstrass (capable of continuation by means of power series) as it presumably is on account of the properties of a regular potential. On this account I have ventured to dedicate this description of subsequent developments to the Gulf Research and Development Company and to Dr. Evjen in particular as his letter⁴⁰ provided the stimulus which led to the development of the method now to be discussed.

6. THE GUIDING EQUATION

The integral equation mentioned in the introduction is

$$f(x) = (1/\pi) \int_{-\infty}^{\infty} \frac{yF(t)dt}{(x-t)^2 + y^2}, \quad (6.1)$$

where y is a fixed positive quantity, $f(x)$ is an analytic function supposed to be determined from observation and presumably of the right type for representation by such an expression, and $F(t)$ is the function to be determined. This function is required to be such that the integral exists and represents a potential $V(x, y)$ regular for $y > 0$ which is the solution of the Dirichlet problem for the half-plane. The properties of this integral of Poisson are well known. To fix ideas we shall suppose that $F(t)$ and $F'(t)$ are finite and continuous for all real values of t , but a case in which $F(t)$ is piece-wise continuous will be considered later.

³⁹ C. Snow, Sci. Pap. Bur. Stand. 19, 231-261, No. 477 (1923).

⁴⁰ M. Evjen, letter dated July 18, 1930.

The assigned values of $f(x)$ may be used to construct a potential

$$W(x, Y) = (1/\pi) \int_{-\infty}^{\infty} \frac{(Y-y)f(t)dy}{(x-t)^2 + (Y-y)^2}, \quad (6.2)$$

which is regular for $Y > y$ and such that $W(x, y) = f(x)$. The potential $W(x, Y)$ is identical with $V(x, Y)$ for $Y > y$; but when $Y < y$, the integral on the right does not represent $V(x, Y)$ but a different potential which has the value $-f(x)$ when $Y = y$. As the particular example

$$\begin{aligned} f(x) &= \cos(mx), \\ W(x, Y) &= e^{-m(Y-y)} \cos(mx) \quad Y > y \\ W(x, Y) &= -e^{m(Y-y)} \cos(mx) \quad Y < y \end{aligned} \quad (6.3)$$

indicates, the derivative $W_y(x, Y)$ is continuous when $Y = y$, and so the correction which must be added to $W(x, Y)$ to give $V(x, Y)$ must have a zero derivative for $Y = y$. Within the region of analyticity of this correction term

$$\begin{aligned} V(x, Y) &= f(x + iY - iy) + f(x - iY + iy) \\ &+ (1/\pi) \int_{-\infty}^{\infty} \frac{(Y-y)f(t)dt}{(x-t)^2 + (Y-y)^2}. \end{aligned} \quad (6.4)$$

Since $V(x, Y) \rightarrow F(x)$ when $Y \rightarrow 0$, we have Snow's inversion formula

$$\begin{aligned} F(x) &= f(x - iy) + f(x + iy) \\ &- (1/\pi) \int_{-\infty}^{\infty} \frac{yf(t)dt}{(x-t)^2 + y^2}, \end{aligned} \quad (6.5)$$

the correctness of which may be checked by direct substitution, if a change in the order of integration is permissible. My friend Stephen O. Rice⁴¹ of the Bell Telephone Laboratories has checked the formula with the aid of Cauchy's theorem.

7. THE CONJUGATE EQUATION

The conjugate of the guiding equation is

$$g(x) = (1/\pi) \int_{-\infty}^{\infty} \frac{(x-t)F(t)dt}{(x-t)^2 + y^2}, \quad F(\pm\infty) = 0. \quad (7.1)$$

The integral on the right represents a potential function $U(x, y)$ regular for $y > 0$ which is the conjugate of $V(x, y)$. In the region $Y > y$, $U(x, Y)$

⁴¹ Stephen O. Rice, letter dated August 30, 1945.

may be represented by Poisson's integral

$$U(x, Y) = (1/\pi) \int_{-\infty}^{\infty} \frac{(Y-y)g(t)dt}{(x-t)^2 + (Y-y)^2}. \quad (7.2)$$

The continuation of this into the region $Y < y$ (or that part in which the added term is analytically continuable) is

$$U(x, Y) = g(x+iY-iy) + g(x-iY+iy) + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(Y-y)g(t)dt}{(x-t)^2 + (Y-y)^2}. \quad (7.3)$$

To find $F(t)$ we write down the conjugate potential

$$V(x, Y) = ig(x+iY-iy) - ig(x-iY+iy) - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(x-t)g(t)dt}{(x-t)^2 + (Y-y)^2}, \quad (7.4)$$

zero being apparently the appropriate value of an additive constant. For $Y=0$ this gives the inversion formula

$$F(x) = ig(x-iy) - ig(x+iy) - (1/\pi) \int_{-\infty}^{\infty} \frac{(x-t)g(t)dt}{(x-t)^2 + y^2}, \quad (7.5)$$

which may be tested by putting $F(t) = \sin(mt)$, $g(x) = -e^{-my} \cos(mx)$. It may be checked by direct substitution and a change in the order of integration.

8. SOLUTION BY MEANS OF ORTHOGONAL FUNCTIONS

When the function $g(x)$ is derived from $f(x)$ by means of the known formula

$$g(x) = (1/\pi)P \int_{-\infty}^{\infty} \frac{f(t)dt}{x-t}, \quad (8.1)$$

in which the P before the integral indicates that the integral has its principle value, a method of solution is possible in which use is made of the expressions for the derivatives of $f(x)$ and $g(x)$

$$f^{(n)}(x) = (n!/\pi y^n) \int_0^\pi F(x+y \cot a) \times \sin(n+1)a \sin^{n-1}a da, \quad (8.2)$$

$$g^{(n)}(x) = -(n!/\pi y^n) \int_0^\pi F(x+y \cot a) \times \cos(n+1)a \sin^{n-1}a da. \quad (8.3)$$

When linear combinations of the functions

$$\begin{aligned} u_n(a) &= \sin(n+1)a \sin^{n-1}a, \\ v_n(a) &= \cos(n+1)a \sin^{n-1}a \end{aligned} \quad (8.4)$$

are used to form the complete set of orthogonal functions,

$$1, \cos 2a, \sin 2a, \cos 4a, \sin 4a, \cos 6a, \sin 6a, \dots, \quad (8.5)$$

it is seen that linear combinations of $f^{(n)}(x)$, $g^{(n)}(x)$ over various values of n will give the Fourier constants of $F(x+y \cot a)$ and so by Fejér's theorem give a practical way of determining this function whenever it is continuous over the infinite range with $F(\pm\infty)=0$. For Fejér's theorem see Whittaker and Watson.⁴²

9. GENERALIZATION OF THE GUIDING EQUATION

Consider a potential $V(x, y)$ defined for the region R outside a curve C by the integral

$$V(x, y) = \int_C G_n[x, y, u(t), v(t)] F(t) dt, \quad (9.1)$$

which solves the problem of Dirichlet for the region R when V has the assigned value $F(t)$ at the point $x=u(t)$, $y=v(t)$ of the curve C . When y has a fixed positive value such that the point (x, y) is outside C for all real values of x , the equation

$$f(x) = \int_C G_n[x, y, u(t), v(t)] F(t) dt \quad (9.2)$$

may be regarded as an integral equation for the determination of $F(t)$. The potential $V(x, Y)$ may be defined for $Y < y$ and for the exterior of C by the expression (6.4) and its analytical continuation. When $x=u(T)$, $y=v(T)$ this gives the inversion formula

$$F(T) = f[u(T)+iv(T)-iy] + f[u(T)-iv(T)+iy] + (1/\pi) \int_{-\infty}^{\infty} \frac{[v(T)-y]f(t)dt}{[u(T)-t]^2 + [v(T)-y]^2}. \quad (9.3)$$

A simple case in which the function G_n is known is that in which C is the circle $x^2+y^2=a^2$. Another case in which G_n is known is that in which the region R lies outside the wedge bounded by

⁴² Whittaker and Watson, *Modern Analysis* (Cambridge University Press, New York), Chap. IX.

the lines $\theta=0$, $\theta=\pi+p\pi$ ($p>0$). It should be noted that in the latter case there are two different paths of analytical continuation of the correction term. One path leads to the upper face of the wedge, the other path leads to the lower face of the wedge. This feature is particularly noticeable when the wedge reduces to a double line $y=0$, $x>0$. Generally the function $f(x)$ has different values at points with equal x but on opposite sides of the line. When, however, the two values of $f(x)$ are equal but of opposite sign, the potential $V(x, y)$ is an odd function of y and is zero for $y=0$, $x<0$. This case will be discussed by another method in Section 1.

10. OBSERVATIONS ALONG A CURVE

Conformal representation may be useful in reducing this case to that already considered but other methods are available. When the curve is the circle $r=a$ (r and θ being polar coordinates), the problem of analytical continuations may be stated as follows: A potential V , regular for $r>a$, has the value $f(\theta)$ at $x=a \cos \theta$, $y=a \sin \theta$ and is given by Poisson's integral

$$V = (1/2\pi) \int_0^{2\pi} \frac{(r^2 - a^2)f(u)du}{r^2 + a^2 - 2(x \cos u + y \sin u)a}. \quad (10.1)$$

For $r<a$ this integral represents the regular potential with the value $-f(\theta)$ at $x=a \cos \theta$, $y=a \sin \theta$. How, then, can a potential V which is regular in a region R including the whole exterior of the circle and part of its interior, including the boundary, be continued from the outside to the inside? The appropriate correction is in this case

$$C(r, \theta) = f[\theta + i \log(r/a)] + f[\theta - i \log(r/a)], \quad (10.2)$$

and is suitable for the solution of the integral equation which arises when $V(x, y)$ gives the solution of the Dirichlet problem for the region outside a curve C lying entirely within the circle $r=a$. When C is a circle, $r=b$; the integral equation and its solution can be easily written down but the result will not be given here because it may be derived by conformal representation from the solution of the guiding equation. Amerio⁴⁸ has recently given a proof, not

⁴⁸ L. Amerio, *Portugaliae Math.* 2, 173-176 (1941).

depending on conformal representation, that a potential assuming analytical values on a segment of an analytic curve can be extended analytically across that curve. The proof depends upon the analytical character of the solution of an integral equation of Fredholm's type which gives the density of the double layer used in the representation of the potential function.

11. GENERALIZATIONS OF THE PROBLEM OF DIRICHLET

Interest in the integral equations of potential theory was much stimulated by the work of Karl Neumann⁴⁴ on the solution by means of series of some integral equations furnished by the problem of Dirichlet and the related problems when use was made of potentials of simple or double layers for the representation of the desired potential. The introduction by Henri Poincaré⁴⁵ of a parameter λ into these problems was an important extension because it led to his surmise that the solution in series could be expressed as the ratio of two entire functions of λ , and this led Fredholm⁴⁶ to seek explicit expressions for these two functions. The work was ably continued by Plemelj⁴⁷ and others and good accounts of the progress along these lines are in books on potential theory such as those of Kellogg⁴⁸ and of Sternberg and Smith.⁴⁹ In the case of Neumann's equation and a vast number of others, the data or assigned values are at points on the line or surface carrying the distribution to be determined. When the data are "free air values," that is, at points not on covered surface, the integral equation is of the first kind, and the observed function is not arbitrary but is restricted by a large number of linear conditions which are often hard to formulate fully. When, however, the data are of the required type, the nature of the type

⁴⁴ K. Neumann, *Akad. Wiss. Wien* 22, 264-326 (1870); *Untersuchungen über das logarithmische und Newton'sche Potential* (B. G. Teubner, Leipzig, 1877).

⁴⁵ H. Poincaré, *Théorie du potentiel Newtonien* (G. Caré et C. Naud, Paris, 1899).

⁴⁶ I. Fredholm, *Acta Math.* 27, 365-390 (1903).

⁴⁷ J. Plemelj, *Monats. Math. Physik* 15, 337-411 (1904); 18, 180-210 (1907).

⁴⁸ O. D. Kellogg, *Potential Theory* (Verlagsbuchhandlung Julius Springer, Berlin, 1929).

⁴⁹ W. J. Sternberg and T. L. Smith, *The Theory of Potential and Spherical Harmonics* (The University of Toronto Press, Toronto, 1944).

of inversion formula to expect is indicated by Snow's series.

$$F(x) = 2 \sum_{n=0}^{\infty} [(-y^2)^n / (2n)!] f^{(2n)}(x) - (1/\pi) \int_{-\infty}^{\infty} \frac{y f(t) dt}{(x-t)^2 + y^2} \quad (11.1)$$

for the solution of the guiding equation. It will be noticed that in addition to an integral involving the observed function $f(x)$ there is also a series involving $f(x)$ and its derivatives. This feature will be useful in the treatment of the problems in three dimensions.

12. THE PROBLEMS IN THREE DIMENSIONS

In space of three dimensions the guiding equation is

$$f(x, y) = (1/2\pi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z F(s, t) ds dt / r^3, \quad (12.1)$$

where

$$r^2 = z^2 + (x-s)^2 + (y-t)^2,$$

and z is a fixed positive quantity. The integral on the right is the well-known solution of the Dirichlet problem for the region $z > 0$. It is the potential of a double layer over the plane and can also be interpreted as the z component of force due to a simple layer on the plane.

For the region $Z > z$ there is a similar integral

$$W(x, y, Z) = (1/2\pi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (Z-z) \times f(s, t) ds dt / R^3, \quad (12.2)$$

where

$$R^2 = (Z-z)^2 + (x-s)^2 + (y-t)^2.$$

For $Z > z$ we have $X(x, y, Z) = V(x, y, Z)$ where V is the potential regular for $Z > 0$ with the value F when $Z=0$. For $Z < z$ the integral W is not equal to V and so by analogy with Snow's formula and our previous method, we introduce a correction in the form of a series and write for $0 \leq Z \leq z$

$$V(x, y, Z) = 2 \sum_{n=0}^{\infty} [(Z-z)^{2n} / (2n)!] \times (-\nabla^2)^n f(x, y) + W(x, y, Z), \quad (12.3)$$

where $\nabla^2 = \partial^2 / \partial x^2 + \partial^2 / \partial y^2$. For $Z=0$ this gives

the inversion formula

$$F(x, y) = 2 \sum_{n=0}^{\infty} [z^{2n} / (2n)!] (-\nabla^2)^n f(x, y) - (1/2\pi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f(s, t) ds dt / r^3. \quad (12.4)$$

When in particular,

$$f(x, y) = \cos(my) f(x), \\ F(x, y) = \cos(my) F(x)$$

there is a simplification and the formula becomes

$$f(x) = (1/\pi) \int_{-\infty}^{\infty} K_1(mr) F(s) ds, \\ r^2 = z^2 + (x-s)^2, \quad (12.5)$$

$$F(x) = 2 \sum_{n=0}^{\infty} [z^{2n} / (2n)!] (m^2 - D^2)^n f(x) - (1/\pi) \int_{-\infty}^{\infty} K_1(mr) f(s) ds, \quad (12.6)$$

where

$$K_0(mu) = \int_0^{\infty} \cos(mt) dt / (u^2 + t^2)^{1/2}, \\ K_1(x) = -DK_0(x), \quad D = d/dx. \quad (12.7)$$

It should be mentioned that S. O. Rice,⁴¹ in the letter already mentioned, has obtained a solution of the equation corresponding to (12.5) with $K_0(mr)$ in place of $K_1(mr)$.

13. THE USE OF LAPLACIAN INTEGRALS

Let us now consider the equation

$$f(x) = (1/\pi) \int_0^{\infty} \frac{y F(u) du}{(x+u)^2 + y^2}, \quad x > 0, \quad y > 0. \quad (13.1)$$

This differs from the guiding equation because (1) the double layer covers only half the x axis, namely the part for which $x < 0$ and (2) because $f(x)$ is assigned only for positive values of x . It should be noticed that the notation is not quite the same as that of Section 6. When a change in the order of integration is valid there is an equivalent formula

$$fx = \int_0^{\infty} F(u) du \int_0^{\infty} e^{-(x+u)t} \sin(yt) dt, \\ = \int_0^{\infty} e^{-xt} \sin(yt) dt \int_0^{\infty} e^{-ut} F(u) du, \quad (13.2)$$

which indicates that (13.1) may be replaced by two integral equations of Laplace's type

$$f(x) = \int_0^\infty e^{-xt} \sin(yt) G(t) dt, \quad (13.3)$$

$$G(t) = \int_0^\infty e^{-ut} F(u) du.$$

If $\sin(yt)G(t)$ and $F(u)$ are required to be continuous, the solutions of these equations, when existent, are unique. Lerch's theory indicates also that a knowledge of $f(x)$ for positive integral values of x or for a set of values in arithmetical progression is sufficient to determine $\sin(yt)G(t)$ and consequently the values of f for non-integral values of x . It should, then, be possible to derive these latter values directly by a formula of interpolation.

14. A FORMULA OF INTERPOLATION FOR LAPLACIAN INTEGRALS

A formula which is particularly suitable for this interpolation has been recommended by the author⁵⁰ and by Erdélyi.⁵¹ It is

$$f(x) = \frac{c_0}{x} + \frac{c_1(1-x)}{x(1+x)} + \frac{c_2(1-x)(2-x)}{x(1+x)(2+x)} + \dots \quad (14.1)$$

The coefficients c_n are found by giving x the values 1, 2, 3, ... in turn, thus $f(1) = c_0$, $f(2) = \frac{1}{2}c_0 - c_1/6$, and it is found that generally

$$c_n = (2n+1)P_n(1-2E)f(1), \quad (14.2)$$

where $P_n(z)$ is the polynomial of Legendre of order n and E is an operator of a distributive nature such that E^s changes $f(1)$ into $f(s+1)$. The series found by expressing each term in the formula of interpolation as a Laplacian integral is in this case

$$(1/\pi) \sin(yt)G(t) = \sum_{n=0}^{\infty} c_n P_n(1-2e^{-t}), \quad (14.3)$$

and this can be used effectively to provide values of $G(t)$ for $t=1, 2, 3, \dots$ with the aid of a table of $P_n(1-2e^{-t})$ for $t=1, 2, 3, \dots$. Table I covers the ranges $n=0(1)10$, $t; 1(1)20$ with 15D, and an attempt is being made to extend it to higher

values of n . The table was calculated from the tables of the exponential function of the Mathematical Tables Project of New York sponsored

TABLE I. $P_n(1-2e^{-t})$.

$t \setminus n$	1	2
1	0.26424 11176 57115	-0.39526 49476 08978
2	.72932 94335 26775	+0.29788 21339 12729
3	.90042 58632 64272	.71615 01028 52814
4	.96336 87222 22532	.89211 89424 35010
5	.98652 41060 01829	.95984 47175 84062
6	.99504 24956 46667	.98516 43522 14122
7	.99817 62360 68891	.99453 36973 78988
8	.99932 90747 44195	.99798 78994 43633
9	.99975 31803 91827	.99925 96325 55358
10	.99990 92001 40475	.99972 76127 88347
11	.99996 65965 98420	.99989 97914 68939
12	.99998 77115 75293	.99996 31349 52388
13	.99999 54793 41186	.99998 50108 31834
14	.99999 83369 42562	.99999 81645 86638
15	.99999 93881 95359	.99999 81645 86638
16	.99999 97749 29651	.99999 93247 89028
17	.99999 99172 01246	.99999 97516 03747
18	.99999 99695 40041	.99999 99086 20123
19	.99999 99887 94407	.99999 99663 83222
20	.99999 99958 77693	.99999 99876 33078
$t \setminus n$	3	4
1	-0.35023 61643 16206	0.13449 18185 76736
2	-0.12412 92757 10654	-0.38184 10855 26292
3	+0.47444 95488 03890	+0.21049 90508 35291
4	.79015 33279 25199	.66302 65464 69307
5	.92050 05158 57439	.86928 43716 52798
6	.97043 89956 51009	.95097 58060 24902
7	.98908 23471 09798	.98183 70921 66184
8	.99597 78237 65385	.99330 08703 23373
9	.99851 95392 12762	.99753 31743 53326
10	.99945 52626 75587	.99909 21868 95476
11	.99979 95879 58828	.99966 59910 88756
12	.99992 62705 84296	.99987 71191 50522
13	.99997 28762 00388	.99995 47938 71677
14	.99999 00216 76114	.99998 33694 87847
15	.99999 63291 74961	.99999 38819 62012
16	.99999 86495 78283	.99999 77492 97645
17	.99999 95032 07525	.99999 91720 12610
18	.99999 98172 40250	.99999 96954 00426
19	.99999 99327 66444	.99999 99879 44074
20	.99999 99752 66157	.99999 99587 76928
$t \setminus n$	5	6
1	0.34415 78146 74583	0.05464 80014 52797
2	-0.40197 48761 18927	-0.21928 12946 63080
3	-0.03838 98178 26693	-0.23878 90479 34199
4	+0.51760 76040 90602	+0.36166 40008 03395
5	.80722 55650 05097	.73556 67348 63206
6	.92691 92142 65318	.89844 75100 59999
7	.97281 77378 76441	.96204 89278 14871
8	.98995 97324 16840	.98595 77710 19097
9	.99630 09031 20756	.99482 31816 62974
10	.99853 09031 20756	.99809 40684 77302
11	.99949 90075 52514	.99929 86457 16643
12	.99981 56815 70533	.99974 19589 36336
13	.99993 21911 90691	.99990 50683 10682
14	.99997 50542 83629	.99996 50760 84201
15	.99999 08229 50036	.99998 71521 41840
16	.99999 66239 47428	.99999 52735 27981
17	.99999 87580 19044	.99999 82612 26878
18	.99999 95431 11656	.99999 93603 40948
19	.99999 98319 16113	.99999 97646 82563
20	.99999 99381 65392	.99999 99134 31550

⁵⁰ H. Bateman, Proc. Nat. Acad. Sci. 25, 262-265 (1939).
 See also E. Hille, Compositio Math. 6, 93-102 (1938).
⁵¹ A. Erdélyi, J. London Math. Soc. 18, 72-77 (1943);
 Phil. Mag. [7] 34, 533-537 (1943).

by the National Bureau of Standards. It is thought that this and similar tables may be useful for the tabulation of functions defined by definite integrals which are hard to compute directly but have simple Laplace transforms.

The convergence of the Legendre series may be improved by a simple device. Thus the equation

$$G(t) = \int_0^\infty e^{-ut} F(u) du \quad (14.4)$$

may be replaced by

$$h(t) = G(t+c) = \int_0^\infty e^{-ut} \cdot e^{-cu} F(u) du, \quad (14.5)$$

and if c is positive the Legendre series for $e^{-cu} F(u)$ derived from $h(t)$ generally converges more rapidly than the series for $F(u)$ derived from

TABLE I (continued).

$t \setminus n$	7	8
1	-0.26817 48023 26647	-0.18068 47715 50211
2	+0.04744 04893 23003	+0.25688 19941 18463
3	-0.36640 21347 36359	-0.40970 70706 13055
4	+0.20339 42382 48305	+0.05093 82503 44370
5	.65573 61016 54870	.56931 56160 02127
6	.86577 13712 05337	.82913 21271 95702
7	.94956 00489 40862	.93538 52422 93312
8	.98129 90110 20097	.97598 81367 27241
9	.99310 05569 44491	.99113 36665 21429
10	.99745 91437 72468	.99673 38012 46168
11	.99906 49156 19968	.99879 78289 73036
12	.99965 59526 47257	.99955 76642 75112
13	.99987 34254 13111	.99983 72627 20031
14	.99995 34349 14456	.99994 01308 03434
15	.99995 28695 40755	.99997 79751 50830
16	.99999 36980 39790	.99999 18974 83377
17	.99999 76816 36173	.99999 70192 47002
18	.99999 91471 21310	.99999 89034 41751
19	.99999 96862 43423	.99999 95965 98696
20	.99999 98845 75400	.99999 98515 96945
$t \setminus n$	9	10
1	0.14819 38218 87370	0.23699 54125 74554
2	0.31162 84080 67464	0.20063 77685 73120
3	-0.37146 40496 78268	-0.26611 53241 48726
4	-0.08810 26036 60170	-0.20710 75794 20539
5	+0.47800 57810 98811	+0.38358 79739 59843
6	.78879 97720 45136	.74507 07436 35930
7	.91956 30995 25272	.90213 67456 08779
8	.97003 04866 29698	.96343 20474 31386
9	.98892 32382 79290	.98647 00901 33615
10	.99591 80856 76644	.99501 21261 41743
11	.99849 73991 57471	.99816 36412 25220
12	.99944 70956 31578	.99932 42487 54669
13	.99979 65804 68941	.99975 13789 98542
14	.99992 51637 84321	.99990 85338 98542
15	.99997 24689 76435	.99996 63510 22666
16	.99998 98718 59350	.99998 76211 68393
17	.99999 62740 59447	.99999 54460 73600
18	.99999 86293 02282	.99999 83247 02917
19	.99999 94957 48383	.99999 93836 92485
20	.99999 98144 96182	.99999 97732 73114

$G(t)$. When the product $e^{-cu} F(u)$ has been derived by means of the Legendre series, the function $F(u)$ may be deduced with the aid of the table of exponential functions already cited.

Equation (14.3) indicates that the series on the right should have zeros at the zeros of $\sin(yt)$. When the series is convergent or summable by the Cesaro method, the equations found by putting $t=1, 2, 3, \dots$ in turn may be regarded as included among the numerous linear conditions that $f(x)$ must satisfy in order that it may be expressible by a definite integral of the form (13.1). The relations derived from the series may be used as a check on the accuracy of the observations or may be used to check the tabular values of a function represented by an integral of type (13.1). In particular, if $F(t)=\pi$, we have $f(x)=\tan^{-1}(y/x)$ and if $y=1$ the coefficients c_n may be calculated with the aid of the table of inverse tangents published by the Mathematical Tables Project mentioned previously. When $t=\pi$ and

$$\mu = 1 - 2e^{-\pi} = 0.91357 \ 21634 \ 72,$$

we have

$$P_2(\mu) = 0.75192 \ 11468, \quad P_3(\mu) = 0.53584, \\ P_4(\mu) = 0.29273 \ 5162.$$

In using such a check it is probably advisable to improve the convergence of the Legendre series by the device already mentioned. The formula of interpolation for Laplace integrals and the table of Legendre functions may be useful in some problems of geophysics in which integral equations of Laplace's type occur. It may be mentioned that one such equation was used by Keck and Colby¹² in working out a method for deriving the depth dependence of earth conductivity from surface potential data.

When the series of Legendre functions is used to solve Eq. (3.3), an equation of type

$$2F(t) = t \sum_{n=0}^{\infty} c_n \int_0^\infty P_n(1-2e^{-z}) J_0(zt) dz \quad (14.6)$$

is obtained, each term of which can be found when the expansion of the Legendre function is used.

¹² W. G. Keck and W. F. Colby, J. App. Phys. 13, 179-188 (1942).

15. SOME PROBLEMS OF ANALYTICAL CONTINUATION

When the solution $F(u)$ of (13.1) has been found, the function $f(x)$ may be continued to negative values of x by means of the formula

$$f(-x) = (1/\pi) \int_0^\infty \frac{yF(u)du}{(x-u)^2 + y^2}, \quad (15.1)$$

the formula of interpolation being unsuitable for this purpose. The function $f(-x)$ can also be found directly by solving the integral equation

$$\begin{aligned} f(x+iy) + f(x-iy) - (1/\pi) \int_0^\infty \frac{yf(u)du}{(x-u)^2 + y^2} \\ = (1/\pi) \int_0^\infty \frac{yf(-u)du}{(x+u)^2 + y^2} \quad 0 < x, \end{aligned} \quad (15.2)$$

which expresses that the function $F(x)$ derived by means of Snow's inversion formula (with an appropriate change of notation) is zero for positive values of x .

It is sometimes possible to continue a function defined by a Laplace integral to negative values of x by means of the equations

$$\begin{aligned} \int_0^\infty e^{-xt} \cos(yt) F(t) dt \\ = \frac{1}{2} \int_0^\infty \cos(xt) e^{-yt} [iF(-it) - iF(it)] dt \\ - \frac{1}{2} \int_0^\infty \sin(xt) e^{-yt} [F(it) + F(-it)] dt, \end{aligned} \quad (15.3)$$

$$\begin{aligned} \int_0^\infty e^{-xt} \sin(yt) F(t) dt \\ = \frac{1}{2} \int_0^\infty \cos(xt) e^{-yt} [F(it) + F(-it)] dt \\ - \frac{1}{2} \int_0^\infty \sin(xt) e^{-yt} [iF(-it) - iF(it)] dt, \end{aligned} \quad (15.4)$$

in which $F(z)$ is supposed to be an analytic function regular when the real part of z is not negative and having a suitable form at infinity. These equations are readily derived by means of Cauchy's theorem. The equation holds for positive values of x and positive values of y but whereas the integrals on the left may diverge for $x < 0$, the integrals on the right may exist and give the desired analytical continuation.

16. POTENTIALS WITH AXIAL SYMMETRY

The equation

$$\begin{aligned} f(z) &= \int_0^\infty [(z+a)^2 + r^2]^{-\frac{1}{2}} F(a) da \\ &= \int_0^\infty e^{-at} J_0(rt) dt \int_0^\infty e^{-at} F(a) da, \quad z > 0 \end{aligned} \quad (16.1)$$

can, when the foregoing change in order of integration is valid, be solved with the aid of two successive inversions of a Laplace integral. For some values of r it may be possible to continue the function $f(z)$ to negative values of z and apply a method of solution to be sketched in Section 17. Some progress in the continuation of Laplacian integrals involving Bessel functions has been made elsewhere by an extension of some work of King⁵³ and Muskat.⁵⁴ The results are included in the general formula (extending that of Pekeris)

$$\begin{aligned} \int_0^\infty e^{-zt} J_0(rt) f(t) dt \\ = (1/\pi) \int_0^\infty K_0(rt) [\cos(zt) \{f(it) + f(-it)\} \\ + \sin(zt) \{f(it) - f(-it)\}] dt, \\ \text{real}(t) > 0, \end{aligned} \quad (16.2)$$

in which $f(t)$ is an analytic function regular in the half plane. This formula and those of Section 15 are included in a more general formula involving Bessel functions as indicated by the author.⁵⁵ Integrals of the present type in which there are no sources above the plane $z=0$ are of interest in aerodynamics, geophysics, and many other subjects. Reference may be made in particular to the work of Theodore von Kármán and others⁵⁶ on the distribution of pressure over an airship in which use is made of a line of sources on the axis of the ship. When the second form of

⁵³ L. V. King, Proc. Roy. Soc. London A139, 237-277 (1933).

⁵⁴ M. Muskat, Physics 4, 129-147 (1933).

⁵⁵ H. Bateman, Monats. Math. Physik 48, 322-328 (1939).

⁵⁶ Theodore von Kármán, Aachen Abhandlungen, No. 6 (1927). See also O. Fuhrmann, *Theoretische und experimentelle Untersuchungen an Ballonmodellen*, Diss. Göttingen (Verlagsbuchhandlung Julius Springer, Berlin, 1912); H. Föttinger, Zeits. f. tech. Physik 9, 26-39 (1928); F. Weinig, Zeits. f. tech. Physik 9, 39-43 (1928); H. Lerbs, Werft, Reederei, Hafen 9, 263-266 (1928).

integral is used with a fairly general type of function $G(t)$ in place of $\int_0^\infty e^{-st} F(a) da$ there may be sources on circles with the z -axis as line of centers.

It may be of interest to form an idea of the equipotential surfaces and streamlines in the whole field of Eulerian flow round the airship though this type of flow is not applicable in the rear of the ship for the flow is useful for finding the actual flow in the boundary layer. We shall suppose, then, that observations are made in front of the ship where the flow is almost exactly of the Eulerian type. If the velocity potential ϕ and stream function ψ are given by

$$\phi = Uz + \int_0^\infty e^{-st} J_0(rt) G(t) dt = Uz + f(z) \quad \text{for a fixed } r, \quad (16.3)$$

$$\psi = \frac{1}{2} Ur^2 - r \int_0^\infty e^{-st} J_1(rt) G(t) dt = \frac{1}{2} Ur^2 + g(z) \quad \text{for a fixed } r, \quad (16.4)$$

we have two Laplacian integral equations for the determination of $G(t)$ when $f(z)$ and $g(z)$ are known. When use is made of the Legendre series for the solution, the derived values of $J_0(rt)G(t)$, $J_1(rt)G(t)$ should be zero at zeros of $J_0(rt)$, $J_1(rt)$, respectively, and so observational values of $f(z)$ and $g(z)$ may be checked.

17. INFINITE LINE OF SOURCES

When sources extend to infinity in both directions, the integral equation

$$f(z) = \int_{-\infty}^\infty [(z-a)^2 + r^2]^{-1} F(a) da \quad (17.1)$$

is harder to solve. It is easy to write down an integral which represents a potential $V(R, z)$ having the value $f(z)$ for $R=r$ and regular for $R>r$. To continue this potential into the region $R<r$ we follow the plan of Section 10 and seek an expansion of V in powers of the quantity $s = \log(R/r)$. Writing

$$V(z, R) = \int_{-\infty}^\infty [(z-a)^2 + r^2 e^{2s}]^{-1} F(a) da, \quad (17.2)$$

the derivatives of V with respect to s can, when $s=0$, be expressed linearly in terms of the

integrals

$$V = \int_{-\infty}^\infty [(z-a)^2 + r^2]^{-n-1} F(a) da, \quad (17.3)$$

and these integrals can likewise be expressed linearly in terms of the derivatives of V with respect to s for $s=0$. Now if

$$g(z) = \int_{-\infty}^\infty \frac{(z-a)F(a)da}{(z-a)^2 + r^2} = V_0 - \frac{1}{2} r^2 V_1 - (r^4/8) V_2 - \dots \quad (17.4)$$

the function $g(z)$ can be regarded as a known function of z when the series converges. The function $F(a)$ can then be derived from $g(z)$ by means of the inversion formula of Section 7.

The expression for V for $R>r$ has the form of an integral involving the function $K_0(mR) = K_0(mre^s)$, the integration being with respect to m . We need, then, an expansion of this function in powers of s . Now if $W = K_0(mre^s)$ we have

$$dW/ds = mre^s K_0'(mre^s),$$

$$d^2W/ds^2 = (mr^2e^s)^2 K_0''(mre^s) + (mre^s) K_0'(mre^s),$$

$$d^n W/ds^n = p_{n-1}(R^2) W + q_{n-2}(R^2) dW/ds,$$

where $p_{n-1}(x)$, $q_{n-2}(x)$ are polynomials of degrees not exceeding $n-1$, $n-2$, respectively. These polynomials satisfy recurrence relations and can be calculated step by step so the n th derivative of W can be found when W and dW/ds are known and it follows that the n th derivative of the potential V for $R=r$ can be found from the values of V and dV/dR for $R=r$.

18. FOURIER ANALYSIS

In 1904 Tedone⁴⁷ found some integral equations of type

$$f(x) = F(x) + (1/2\pi) \int_{-\infty}^\infty k(t) dt \times \int_{-\infty}^\infty \cos[t(x-u)] F(u) du \quad (18.1)$$

in a discussion of the equilibrium of an infinitely long elastic cylinder with given surface displacements. With the abbreviation $C = \cos t(x-u)$ the

⁴⁷ O. Tedone, Acc. dei Lincei, [5] 13, 232-240 (1940).

solution used was

$$F(x) = (1/\pi) \int_{-\infty}^{\infty} H(t^2) dt \int_{-\infty}^{\infty} C \cdot f(u) du, \quad (18.2)$$

where

$$H(t^2) = \frac{1}{2 + k(t) + k(-t)}. \quad (18.3)$$

With Snow's symbolical method

$$F(x) = 2H(-D^2)f(x), \quad \text{where } D = d/dx. \quad (18.4)$$

In the case $f(x) = M \cdot g(x)$, $k(t) = M \cdot h(t)$, where M is a large positive number, the limit $M \rightarrow \infty$ gives the solution of an integral equation of the first kind⁵⁸

$$g(x) = (1/2\pi) \int_{-\infty}^{\infty} h(t) dt \int_{-\infty}^{\infty} C \cdot F(u) du, \quad (18.5)$$

$$F(x) = (1/\pi) \int_{-\infty}^{\infty} \frac{dt}{h(t) + h(-t)} \int_{-\infty}^{\infty} C \cdot g(u) du. \quad (18.6)$$

In particular, if

$$h(t) = e^{-\nu|t|}, \quad (18.7)$$

the equation becomes the guiding equation of Section 6,

$$g(x) = (1/\pi) \int_{-\infty}^{\infty} \frac{yF(u)du}{(x-u)^2 + y^2}, \quad (18.8)$$

and the solution is found in the form

$$F(x) = (1/\pi) \int_0^{\infty} e^{\nu|t|} dt \int_{-\infty}^{\infty} C \cdot f(u) du \quad (18.9)$$

given by Snow and Evjen. When, moreover, use is made of the identity

$$e^{\nu|t|} = 2ch(yt) - e^{-\nu|t|},$$

the integral involving the hyperbolic cosine may be evaluated by the symbolic method and the solution obtained in the form of a simple integral with additional terms.

It must be emphasized that when use is made of the Fourier analysis the function $F(u)$ must be restricted more, perhaps, than when use is made of Poisson's integral directly. Evjen, for instance, required

$$\int_a^{\infty} |F(u)| du \quad \text{and} \quad \int_{-\infty}^{-a} |F(u)| du$$

⁵⁸ When this result was sent to S. O. Rice, he remarked in a letter (reference 41) that there is a slight simplification when the exponential form of Fourier's integral is used.

to be convergent but other restrictions are permissible when Lebesgue integrals are used.

19. AN INTEGRAL EQUATION OF CRYSTAL THEORY

When M is kept finite and $k(t)$ has the same form as before, Tedone's equation becomes

$$f(x) = F(x) + (M/\pi) \int_{-\infty}^{\infty} \frac{yF(u)du}{(x-u)^2 + y^2}, \quad (19.1)$$

and his solution takes the form

$$F(x) = (1/2\pi) \int_{-\infty}^{\infty} \frac{dt}{1 + Me^{-\nu|t|}} \int_{-\infty}^{\infty} C \cdot f(u) du. \quad (19.2)$$

The integral equation in this case includes one used by Hulthén⁵⁹ in work on the exchange problem for crystals. To get a solution more suitable for computation we write

$$\frac{1}{1 + Me^{-\nu|t|}} = 1 - \frac{Me^{-\nu|t|}}{1 + Me^{-\nu|t|}}, \quad (19.3)$$

$$F(x) = f(x) - M \int_{-\infty}^{\infty} G(M, x-u, y) f(u) du, \quad (19.4)$$

where

$$G(M, x, y) = (1/\pi) \int_0^{\infty} \frac{e^{-\nu t}}{1 + Me^{-\nu t}} \cos(xt) dt \quad y > 0 \quad (19.5)$$

$$= (1/\pi y) \int_0^{\infty} \frac{e^{-\nu v}}{1 + Me^{-\nu v}} \cos(xv/y) dv.$$

For sufficiently small values of (x/y) there is an expansion

$$G(M, x, y) = (1/M\pi y) [\phi(1, -M) - (x/y)^2 \phi(3, -M) + (x/y)^4 \phi(5, -M) - \dots], \quad (19.6)$$

where

$$\phi(s, z) = \sum_{n=1}^{\infty} n^{-s} z^n \quad (19.7)$$

is the function studied by Appell,⁶⁰ Jonquière,⁶¹ Truesdell,⁶² and others. The function has been used in geophysical work by King.⁶³

⁵⁹ L. Hulthén, Arkiv Mat. Astronom. Fysik A26, No. 11 (1938).

⁶⁰ P. Appell, Comptes rendus 87, 690-692, 874-878 (1878).

⁶¹ A. Jonquière, Bull. Soc. Math. France [2] 17, 142-152 (1889), Stockholm Acad. Öfversigt 257-268 (1889).

⁶² A. Truesdell, Ann. of Math. 46, 144-157 (1945).

Here and There

New Appointments

Dr. William A. Shurcliff of Washington has been appointed technical consultant in the Bureau of Industry of the New York State Department of Commerce.

Edmund S. Winlund is the new RCA Victor Industrial Electronics Engineer for the Pacific Region, with headquarters in Los Angeles.

C. W. Handley and P. D. Ries have been selected as technical specialists by National Carbon Company, Inc., to help expand the company's technical service facilities applying to projector carbons and other lighting carbon products.

Dr. Emilio Segrè has been appointed professor of physics at the University of California at Berkeley.

Drs. Dana P. Mitchell and Norman F. Ramsey, Jr., of Columbia University, have been promoted to the rank of associate professor of physics.

Dr. Louis B. Slichter, professor of geophysics, Massachusetts Institute of Technology, will become professor of geophysics, University of Wisconsin, with the opening of the second semester in January.

Five new vice presidents have been announced by RCA Victor Division of Radio Corporation of America: Meade Brunet, J. B. Elliott, Joseph H. McConnell, J. W. Murray, and L. W. Teegarden.

Dr. John A. Tiedeman, formerly associate professor of physics at Woman's College, University of North Carolina, has been appointed director of the Education Department of Ansco Division of General Aniline and Film Corporation.

Professor Marcel K. Newman has been appointed Associate Professor of Mechanical Engineering in the College of Applied Science, Syracuse University.

Dr. Lise Meitner, the Austrian physicist, has joined this month the faculty of the Catholic University of America, where she is visiting professor.

Back to Former Posts

Joseph Kaplan, professor of physics, University of California at Los Angeles, has returned to his post after five years of service in the Army Air Forces.

Dr. Robert S. Shankland, on leave for the past three years to the Office of Scientific Research and Development, has returned to his position as head of the Department of Physics at Case School of Applied Science, and Drs. Robert Rinehart, Charles C. Torrance, Paul Guenther, and

Lewis J. Green have resumed their work in the Department of Mathematics at Case.

National Academy Appointment

Science reports that Dr. Detlev W. Bronk, professor of biophysics and director of the Eldridge R. Johnson Foundation of the University of Pennsylvania, has been appointed foreign secretary of the National Academy of Sciences. By virtue of this appointment, Dr. Bronk also becomes chairman of the Division of Foreign Relations of the National Research Council.

Michigan Society of Spectroscopy

The Michigan Society of Spectroscopy has elected new officers for the second year of its existence. New chairman is G. A. Nahstoll of the Ford Motor Company; Secretary is Virginia Sink of the Chrysler Corporation.

Nuclear Physics Program at Harvard

The president and fellows of Harvard College have voted to establish a committee to develop a program in nuclear physics at the university. The corporation has further voted to allocate the sum of \$400,000 to the committee to spend within a period of five years in developing the program.

Lighting Handbook

Plans for publication of a lighting handbook, most comprehensive and informative work of its kind ever attempted, have been announced by the Illuminating Engineering Society. The handbook, long contemplated by the Society as a major post-war project, will contain five hundred pages of text prepared by outstanding authorities in their respective fields. Every phase of lighting, from the pure physics of light to specific lighting recommendations for stores, offices, homes, factories, and even for juke boxes and television studios, will be covered. The latest authoritative information on light sources, as well as on the measurement and control of light, will be included. The handbook will be released October, 1946 and will be sold and distributed through official I.E.S. channels. Robert W. McKinley, electrical engineer, has been engaged as editor.

Necrology

Dr. Eugene Cook Bingham, professor of chemistry at Lafayette College, died on November 6 at the age of 56.

National Academy Awards

Among the awards given by the National Academy of Sciences on November 16 were the following:

Cyrus B. Comstock Prize, awarded every five years for the most important discovery or investigation in electricity, magnetism, and radiant energy. Prize of \$3,000 for the period 1938 to 1943 was awarded to Donald W. Kerst, Department of Physics, University of Illinois, "for his pioneer work in connection with the development of the betatron and the results which he has obtained with this new and powerful scientific tool."

The gold Public Welfare Medal, for eminence in the application of science to the public welfare, was awarded for 1945 to Vannevar Bush, Director of the Office of Scientific Research and Development, Washington, D. C., "in recognition of his outstanding service in bringing to bear the scientific and engineering talent of this country upon problems of research connected with the war effort."

Nobel Prizes

The 1945 Nobel prize in physics has been awarded to Professor Wolfgang Pauli of the Institute for Advanced Study at Princeton, New Jersey, for his work on the "exclusion principle," which deals with regulation of electrons in the outer shell of atoms and molecules. The 1944 prize in chemistry has been awarded to Professor Otto Hahn, Germany, an authority on radioactivity, for his discovery of a method of breaking the heavy atom nucleus. Artturi Wirtanen, of Finland, biochemist, has received the 1945 award in chemistry for his discoveries relating to food chemistry.

Atomic Energy Forum at Cleveland

On November 27 the Cleveland Physics Society, Case School of Applied Science, and Western Reserve University sponsored a forum on the control of atomic energy. It was attended by more than 500 scientists in the Cleveland area. Main speaker was Dr. W. H. Zinn, nuclear physicist from the University of Chicago. At the end of the meeting the following resolution was passed almost unanimously by the audience:

"RESOLVED, that the Government of the United States be urged to take rapid, effective steps to make the United Nations Atomic Energy Commission a reality. Further, that all effort be made to secure international cooperation resulting in free exchange of scientific information among all countries and unhampered inspection by United Nations experts to insure that such information is not misused. Also, that the Government of the United States be urged to take such legislative steps as will avoid compromising either atomic or nuclear research in this country or the establishment of an adequate international policy. Further, that this resolution be submitted to all government officials immediately concerned.

Revised Fulbright Bill

Hearings on the revised Fulbright Bill, S. 1248, to establish an Office of Technical Services within the Department of Commerce, consolidating several similar or related scientific agencies of the Government, began December 12. The revised bill had been drafted to conform to recommendations of various agencies of the Government and private organizations that might be affected by or interested in the proposed legislation. Among these were the Departments of Commerce and Agriculture, the Office of Scientific Research and Development, the National Inventors Council, the Smaller War Plants Corporation, and Office of Production Research and Development.

Crystallographic Society

The Crystallographic Society organized at Harvard University and Massachusetts Institute of Technology in 1939 is planning to resume its activities, which were suspended during the war. The Society concerns itself with the science

of crystallography and its applications to fields such as physics, chemistry, metallurgy, ceramics, and biology and is not restricted to the classical phases of crystallography commonly associated with mineralogy. In 1940 the Society initiated the idea of launching a Journal of Crystallography. More recently, plans have been discussed by several societies for starting a new international Journal of Crystallography to replace the Zeitschrift für Kristallographie, which is no longer published. Since there is no other journal devoted exclusively to crystallography, the Society takes an active interest in this new development.

A meeting to be held at Smith College is tentatively planned for March 21-23, 1946. Those wishing to present papers are requested to send title and brief abstract to Professor M. J. Buerger, Massachusetts Institute of Technology, Cambridge 39, Massachusetts. Ten invited papers have been arranged.

All those wishing to renew membership or to become members of the Society or who desire further information may write to William Parrish, Acting Secretary-Treasurer, The Crystallographic Society, Philips Laboratories, Inc., Box 39, Irvington, New York. Membership fees have been set at \$1 for 1946. A new membership list will be compiled January 31, 1946.

Society for Quality Control

On October 3 and 4, 1945, the Organizing Committee for a National Quality Control Society met at Pittsburgh, Pennsylvania, and established the Society for Quality Control. It is the intent of the Society "to serve all those interested in quality control, giving particular attention to the needs of those engaged in manufacture and inspection." The Society plans "to publish material that will be of help to quality control people everywhere, material that will help them sell and apply quality control methods."

Edward M. Schrock and Ralph E. Wareham were unanimously elected president and secretary-treasurer, respectively. Dues of the Society were set at \$5, of which \$3 will be returned to the local chapter. At present there are fourteen chapters. Payment of dues to the Society will make a person a member of the chapter in his geographical area. Dues should be sent to Mr. Wareham at 305 East 43d Street, New York 17, New York.

Dr. W. D. Coolidge Honored by Chile

Dr. William D. Coolidge, formerly vice president and director of research of the General Electric Company, who is now touring South America, has been awarded the "Orden al Merito" of the Chilean Government.

Division of High-Polymer Physics

The Division of High-Polymer Physics of the American Physical Society held its second regular meeting at Columbia University, New York City, January 24-26, 1946, as one of the features of the general, annual meeting of the parent society. On the program of the Division meeting were approximately 25 papers on rubbers, plastics, and other high-polymeric materials, covering both the analysis of physical behavior and experimental techniques.

The Bulletin of Mathematical Biophysics

#

MARCH, 1946

The Mathematical Biophysics of Some Mental Phenomena: II. Anxieties and Elations—N. RASHEVSKY.

Chain Processes and Their Biophysical Applications: Part II. The Effect of Recovery—I. OPATOWSKI

A Contribution to the Mathematical Biophysics of Cell Growth and Shapes: II.—RICHARD RUNGE

A Theory of Membrane Permeability: II. Diffusion in the Presence of Water-Flow—INGRAM BLOCH

The Neural Mechanism of Logical Thinking—N. RASHEVSKY

Volume 8, Number 1—The University of Chicago Press, Chicago, Illinois.

Fellowships in Cancer Research

Fellowships in cancer research, administered by the Committee on Growth of the National Research Council, are open to citizens of the United States who possess M.D. or Ph.D. degrees. They are designed to provide opportunities for training and experience in research for men and women who are fitted for scientific investigation, basic or clinical (including research in biophysics). The fellowships are intended for recent graduates who, as a rule, are not more than thirty years of age.

Applicants for these fellowships may plan to enter any field of research pertaining broadly to the problem of cancer. Candidates will be favored who desire to obtain thorough experience in one of the basic sciences.

The annual stipends are determined by individual circumstances and by cost of living in the location of study. The usual amount is from \$2000 to \$3000 per annum. The fellowships are granted for one year, but they may be renewed. Fellows are usually chosen at annual meetings of the Committee on Growth in February, and applications to receive consideration at these meetings must be filed on or before January 1. During 1945-46, interim applications will be considered. Appointments may begin on any date determined by the Committee.

Further particulars concerning these fellowships may be obtained on request from Executive Secretary of the Committee on Growth, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

Radiation Laboratory Technical Series

Publication of the largest series of books on physics and electrical engineering ever undertaken in the United States has been arranged by the Massachusetts Institute of Technology for the Office of Scientific Research and Development, according to an announcement by James R. Killian, Executive Vice President of the Institute. The Radiation Laboratory Technical Series, comprising twenty-eight titles and a General Index, reports the results of five years' wartime work on radar. The purpose of the series is to make available to science, industry, and the public generally the results of the immense developments in electronics and in microwave theory and technique during the war years, which, once secret, have now widely been declassified. These will be of inestimable value in peacetime research in physics, biology, and other natural sciences, as well as providing the engineering foundation for

post-war industrial developments in television, communications, and electronics.

Work reported in the series required the expenditure of about 20,000 technical man-years in the various radar research and development establishments, Dr. Killian revealed. "For the first time," he said, "the technical literature of a large subject is being created all at once, on a uniform basis. Emphasis in the series will not be on radar itself, but rather on the basic techniques which underlie many phases of electronics in addition to radar."

Publication of the series will be handled by an agreement between Massachusetts Institute of Technology, acting for the Office of Scientific Research and Development, and the McGraw-Hill Book Company, which is undertaking the publication. This arrangement is expected to serve as a model for the publication of other scientific work done with public funds under OSRD contract, now that the need for wartime secrecy no longer exists where this information is of general interest and can desirably appear in book form. Royalties on the sale of the books will be paid to the U. S. Treasury.

Plans for the series have been under way since late 1944, when it became clear that the end of the war was not far off. Intensive work has been carried on since June. Manuscript is now being prepared at the rate of nearly 400 pages per day by 187 authors under the direction of nine editors. Illustrations are being drawn by more than a hundred artists and draftsmen.

Books of the series are being written chiefly by staff members of the Radiation Laboratory, wartime radar research center maintained at the Massachusetts Institute of Technology by OSRD contract, but will include the results of work on radar done in British development establishments and in industrial laboratories both here and in England. Several British scientists have come to the Massachusetts Institute of Technology to cooperate in the preparation of the series.

Summaries of the scope and contents of the series show that it covers several fields of the greatest scientific and engineering importance. Some of these are: precise timing techniques; new methods of cathode-ray tube display; generation, transmission, and radiation of high-power microwaves; and broad-band amplifier techniques. Present titles planned for the series follow:

The generation of high power microwaves is dealt with in a book on *Microwave Magnetrons* which presents the theory of operation of these oscillators as well as practical design considerations and operating techniques. *Low Power Microwave Tubes* deals principally with the properties of reflex klystrons and lighthouse tubes when used as oscillators, amplifiers, rectifiers, and mixers. Production of accurately timed pulses having various waveforms at high and very high peak powers is treated in *Pulse Generators*. A theoretical and practical treatment of pulse transformers is also included in this book.

Transmission line and wave guide techniques are discussed exhaustively. *The Theory of Guided Waves* develops a basic theory of electromagnetic waves in wave guide. This theory draws from conventional field theory, circuit theory, and transmission line theory, but is directed to the solution

of problems of importance in the microwave region. Maximum use of engineering concepts is made. A *Waveguide Handbook* collects the analytical results given by the theory and presents all available numerical and experimental results in graphic form, chosen to be most convenient for practical circuit design. *The Principles of Microwave Circuits* develops a generalization of low frequency network theory and of the impedance concept. The properties of wave guide circuit elements and devices are developed and discussed in terms of this generalization. *Microwave Transmission Circuits* treats from a practical point of view the same general matters. Design principles for connectors, rotary joints, and other wave guide and transmission line devices will be discussed, and examples of successful designs given. *The Techniques of Microwave Measurements* describes in detail the methods for measuring power and attenuation at high and low level, standing wave measurements, means for accurate determination of wave-length and frequency, r-f spectrum and pulse shape, and other means of measurement peculiar to the microwave field.

Microwave Antenna Theory and Design provides a comprehensive survey of theory and design techniques for microwave antennas, a full discussion of antenna measurement methods, and an indication of special methods used in antenna production. *Propagation of Short Radio Waves in the Troposphere* collects and summarizes the very extensive wartime investigations of the propagation characteristics of radiation at frequencies too high to be affected by the ionosphere. The theory of atmospheric refraction, the meteorology of the refraction problem, and the experimental approach to the refraction problem will be followed by a treatment of matters which include target properties, ground and sea clutter, and molecular absorption.

Crystal Rectifiers discusses the theory, properties, manufacture, and use of the silicon and germanium point contact rectifiers which have been developed for use as microwave converters and other circuit applications. *Microwave Receiving Circuits* deals with the problem of frequency conversion and duplexing. Means of automatic frequency control for a local oscillator are discussed, and designs of mixers and duplexing assemblies are described.

A series of works on the design and application of vacuum tube circuits is opened with the *Components Handbook* which describes the properties of available wire, resistors, capacitors, inductors, instrument motors, potentiometers, and other circuit elements. Emphasis is placed on the specifications which components meet and on features of performance which are not usually given in other publications. The book includes the results of independent measurements of important properties of various types of commercial components. *Cathode Ray Tubes* describes methods for using such tubes, with emphasis on good focus, freedom from distortion, and reliable operation. It includes discussions of focusing and deflection magnets and coils, the properties of fluorescent screens, and methods of constructing auxiliary apparatus such as projectors, magnetic shields, light filters, and the like.

Two books follow which analyze the fundamental properties of basic circuits useful in the various branches of electronics. *Vacuum Tube Amplifiers* deals with circuits

which can be treated theoretically by linear circuit analysis. It seeks to analyze completely many types of amplifiers, especially those of high gain, wide band pass, or large dynamic range. Constructional details of special amplifiers, and rigorous critical discussions of amplifier sensitivity will be given. *Waveforms* discusses basic circuits which include a non-linear element. These and other circuits for the generation and shaping of the form of current and voltage waves are analyzed with the help of the cathode-ray oscillograph. The wave forms described make possible radar and television displays, and permit the precise measurement of time and position. A treatment of their generation is followed by discussions of their manipulations, such as modulation, analysis, and frequency division.

Then follow those books whose object it is to present design techniques for synthesizing from the basic circuits and components just described the complex instruments required to fill various functional demands. *Electronic Instruments* deals with devices for the purpose of precision time measurement, data transmission, and mathematical computation. It emphasizes instrument function, gives details of engineered circuit designs, and recommends preferred equipment types where possible. *Cathode Ray Tube Display Circuits* shows how cathode-ray tubes may be combined with electronic circuits to provide a wide variety of measuring and precision data display devices. Applications of such devices include the precision measurement of time intervals shorter than one ten-thousandth of a microsecond, and the plotting of several functions of several variables simultaneously in any of several coordinate systems, as well as special devices suitable for radar and television use.

The use of electrical and other time-variable indications in automatic control devices is discussed in *Automatic Control Systems*. Basic principles for the design of electrical and mechanical feedback control systems are developed in detail, and application is then made to a series of automatic control problems, including those which have arisen in automatic radar range and angle tracking.

Microwave Receivers describes many different types of complete receiving systems, suitable for radar, television, relay telephony, and repeat-back devices. Examples are chosen to illustrate actual design techniques. *Signal Thresholds in Interference* offers an analysis, both theoretical and experimental, of the factors affecting the perception of desired signals in the presence of various types of interference, principally receiver noise.

An outline of the general principles of design of radar systems is presented in a volume entitled, *Radar Systems Engineering*. It is intended as a basic treatise and reference book for anyone interested in making any application of radar.

The applications of radar to problems of air and sea navigation, including airport and harbor traffic control, are discussed in a volume entitled, *Radar Aids to Navigation*.

In the book, *Loran*, a comprehensive treatment is made of the principles and engineering design of this war-born long range navigational aid.

The final book is a general index providing cross references among the various books of the technical series.

Symposium on Molecular Spectroscopy and Molecular Structure

A conference on Industrial and Chemical Infra-red Spectroscopy will be held at Mendenhall Laboratory of Physics, Ohio State University in Columbus, Ohio, June 13-15, 1946. This meeting will be sponsored jointly by the Physics departments of the University of Michigan and the Ohio State University.

The conference will be a part of a larger symposium on Molecular Spectroscopy and Molecular Structure to be held at the Ohio State University beginning Monday, June 10 and extending to June 15.

More complete details on this symposium will be published in a later issue of the Journal. Inquiries concerning the infra-red conference may be directed to Professor Harald H. Nielsen, Mendenhall Laboratory or to Dr. Norman Wright, Dow Chemical Company, Midland, Michigan.

Naval Ordnance Laboratory

The Charles H. Tompkins Company, of Washington, has begun work in preparation for the construction of the main buildings of the Naval Ordnance Laboratory's new research center at White Oak, Maryland. The contract, totaling \$4,813,000 was awarded the Tompkins Company for construction of a group of structures which will include the administration building, the photographic laboratory, spherical field laboratory, long field laboratory, officers' quarters, Marine barracks, and a portion of the roads which will connect the numerous units in the 938-acre area at White Oak.

A total of 50 permanent buildings is planned for the research center which the Naval Ordnance Laboratory will use for the peacetime continuance of scientific experiments in connection with underwater ordnance. More than a dozen of the smaller buildings have already been completed, and a \$300,000 boiler plant—which will supply steam for heating and for process work in the main laboratory—will be ready within a few weeks. Cost of the entire project will be approximately \$15,000,000.

Included in the buildings now in operation are six magnetic laboratories, which are located in a remote section of the White Oak property away from all local media of interference. They contain no properties which would tend to create disturbances during experiments, and are devoid of red brick—which contains iron oxide—and reinforced concrete—which requires the use of iron rods. Downspouts, radiators, and nails are all of copper.

One of these magnetic buildings, the Quiet Laboratory, has no electric lights, water pipes, or other ordinary facilities which could in any possible way interfere with the work of the men who will be probing into the mysteries of magnetic weapons and counterweapons. Another structure in this group has a temperature control room for producing immediate temperature variations of between minus 80 and plus 180 degrees Fahrenheit.

The new Laboratory was planned by the Navy Department's Bureau of Ordnance in collaboration with Captain W. G. Schindler, USN, Officer-in-Charge of the Naval Ord-

nance Laboratory, and Captain R. D. Bennett, USNR, its technical director. Captain Schindler came to the Laboratory's present headquarters, at the Navy Yard in Washington, two years ago after a brilliant career as Ordnance Officer in a carrier squadron in the South Pacific. He made numerous sorties as rear seat gunner in an SBD dive bomber, and is credited with being the first American to shoot down a Japanese plane in aerial combat. Captain Bennett, internationally known for the development of cosmic-ray measuring equipment, is a former professor of electrical engineering at Massachusetts Institute of Technology.

Most of the construction details for the gigantic laboratory were handled through the Bureau of Yards and Docks, but in NOL's own offices were prepared the unique laboratory features designed to provide the technical staff with the world's most modern equipment for the development of the various projects which the Laboratory will inherit.

Architects are Eggers and Higgins, of New York City, with Taylor and Fisher, Baltimore, associates.

Naval Ordnance Laboratory has long been the Navy's headquarters for the design and development of underwater weapons and countermeasures. It was established in 1918 as a mine laboratory, but because of the country's "peace economy" program only a skeleton staff was retained for the next two decades. When Germany introduced the magnetic mine at the beginning of World War II, the Laboratory immediately began to expand its activities to include the development of depth charges, bomb fuzes, projectile fuzes, pyrotechnics, and numerous other items as well as mines. Perfection of degaussing equipment, which was to protect our ships from lurking enemy magnetic mines, was one of the first tasks assumed by the Laboratory after the dark months of 1940.

Until the last days of Japan's resistance, most of the Laboratory's work was restricted from public attention because of security regulations. Announcement of the mining of the Inland Sea and other strategic Japanese waters first turned the spotlight on NOL, and finally the Navy announced that it was this research center that had developed all the mines which were used in the history-making blockade. In addition, the Laboratory had perfected the parachutes which eased the deadly missiles into the water from an altitude of 10,000 feet.

Another of the Naval Ordnance Laboratory's outstanding contributions which was withheld from the public until the cessation of hostilities was the development of the 40-mm fuze. Through this device the Navy's guns were able to destroy enemy aircraft with greater accuracy, and at the same time the U. S. taxpayer was given considerable help in the cost of producing the ammunition. NOL's technical experts are credited with having effected a saving of \$236,820,000 in the three years that the new fuzes were in production.

More than a thousand scientists and technicians will make up the peacetime staff of the Laboratory. Nearly one hundred of the scientific personnel have already moved to the new workshops at White Oak, with others following as the various buildings are completed. The entire staff is expected to be transferred by the late summer of 1946.

Federation of Atomic Scientists

The following statement has been prepared regarding the aims of the newly formed Federation of Atomic Scientists:

We, the undersigned representatives of the Associations of Scientists who have worked on the atomic bomb, hereby agree to form a united group, to be known as "The Federation of Atomic Scientists," in order to carry out more effectively the common aims of the separate organizations.

Each of the six Associations shall retain its identity and independence of action. The Federation will provide a central office and staff for the purpose of aiding and coordinating the activities of the several member Associations.

The component organizations were founded to achieve the following aims:

- (1) To study the implications to our nation and to the world of the liberation of nuclear energy.
- (2) To create a realization of the dangers that this nation and all civilization will face if the tremendous destructive potential of nuclear energy is misused.
- (3) To help establish an atmosphere of world security in which the beneficial possibilities of nuclear energy may be developed.
- (4) To study the relation between national legislation and the establishment of an adequate international policy.

. . . and to give all possible publicity to the following convictions:

- (1) That a continuing monopoly of the atomic bomb by the United States is impossible.
- (2) That there can be no specific defenses against the destructive effects of the atomic bomb.

- (3) That in view of the existence of the atomic bomb, no nation can, in this new age, feel secure until the problem of the control of atomic power is solved on a world level.

The council of the Federation will consist of those delegates of the component associations who are in Washington at any given time. It is intended that one such member from each association should be present in Washington at all times and two will frequently be present. There will be a central office which will act primarily as headquarters for the Associations. It will also serve as an information and speakers' bureau and will handle contacts with other groups which hold views similar to our own.

The Washington office shall be made available to all scientists' organizations in America which find it necessary to have the same information that we are to supply to the Associations. Many of these newly formed groups have the same aims and purposes as our own organization.

Signed by representatives from:

Association of Oak Ridge Scientists at Clinton Laboratories
The Atomic Scientists of Chicago
The Association of Los Alamos Scientists
The Association of Manhattan Project Scientists, New York City Area
The Atomic Production Scientists, Oak Ridge
The Atomic Engineers, Oak Ridge

The Federation of Atomic Scientists may be reached by calling National 5818, Washington, D. C. Its street address is 1621 K Street, N.W., Washington 6, D. C.

New Booklets

National Patent Council, new organization devoted to the interests of small manufacturers, is issuing an 8-page monthly bulletin called *Competitor*. It is published at Gary, Indiana. The December issue reported progress to date in enrolling new support for the campaign.

The National Research Bureau, Inc., 415 North Dearborn Street, Chicago 10, Illinois has issued for distribution a list of some of the engineering and manufacturing problems which have been solved by the Bureau. Correspondents may order copies of the Bureau's reports on any listed problems. The Bureau also offers a list of new Masters' theses which are available for distribution. Such theses will be supplied only on microfilm.

American Instrument Company, 8030 Georgia Avenue, Silver Spring, Maryland, publishes an 8-page bi-monthly paper entitled *Aminco Laboratory News*, which is "devoted to the application of scientific instruments to modern laboratory technique, engineering, materials testing, and production control."

The Senate Subcommittee on War Mobilization has issued a monograph on *Wartime Technological Developments*, a supplement to the previous volume of the same name. Like that volume, this monograph was compiled and digested by the Bureau of Labor Statistics, U. S. Department of Labor, from items published in trade and technical journals. There are 839 technical advances included in the volume.

Calendar of Meetings

February		18-20	American Philosophical Society, Philadelphia, Pennsylvania
23	American Mathematical Society, New York, New York	25-27	American Physical Society, Cambridge, Massachusetts
25-28	American Institute of Mining and Metallurgical Engineers, Chicago, Illinois	26-27	American Mathematical Society, Chicago, Illinois
25-28	Technical Association of the Pulp and Paper Industry, New York, New York	26-27	American Mathematical Society, New York, New York
25-March 1	American Society for Testing Materials, Pittsburgh, Pennsylvania	27	American Mathematical Society, West Coast
		28-May 1	American Ceramic Society, Inc., Buffalo, New York
March		May	
7-9	Optical Society of America, Cleveland, Ohio	6-10	Society of Motion Picture Engineers, New York, New York
27-30	American Association for the Advancement of Science, St. Louis, Missouri	10-11	Acoustical Society of America, New York, New York
		27-29	American Geophysical Union, Washington, D. C.
April		June	
1-4	American Society of Mechanical Engineers, Chattanooga, Tennessee	17-20	American Society of Mechanical Engineers, Detroit, Michigan
3-5	Society of Automotive Engineers, New York, New York	20-23	Society for the Promotion of Engineering Education, St. Louis, Missouri
8-12	American Chemical Society, Atlantic City, New Jersey	24-28	American Institute of Electrical Engineers, Detroit, Michigan
12-13	American Physical Society, Southeastern Section, Atlanta, Georgia	24-28	American Society for Testing Materials, Buffalo, New York
17-19	American Society of Civil Engineers, Philadelphia, Pennsylvania		

Errata

On the Representation of Rigid Rotations

[J. App. Phys. 16, 571 (October 1945)]

H. SCHWERTFEGER

THE last line of footnote 2 must read: "*scientifiques* 649 (Paris 1938)."
In line 15 of section 2, two different symbols are used for the exponential function ($\exp P$ and e^P) in one and the same formula.

On Maximum Gain-Band Width Product in Amplifiers

[J. App. Phys. 16, 528 (September 1945)]

W. W. HANSEN

DR. Henry Wallman of M.I.T. has pointed out that there are certain contradictions in this article. Specifically, in the right column of page 529, one finds "Second, Wheeler's four-terminal result is also the best possible for a single stage." This statement, which occurs also in the conclusion, conflicts with the correct result of Bode, which is given in the abstract and on page 533.

The opportunity for this carelessness on my part arose because I did not see Bode's important work until my article was in press and, while modifying some of my statements, I failed to correct them all.

Calculation of the Output from Non-Linear Mixers

HARRY STOCKMAN

Cruft Laboratory, Harvard University, Cambridge, Massachusetts

(Received October 6, 1945)

The exact mathematical treatment of non-linear mixers of the type used in very-high frequency and ultra-high frequency superheterodynes is difficult. If, however, the non-linear element is simplified to the extent that it is fully described by a plotted current-voltage characteristic, several ways of determining the output become evident. Some of the most fundamental methods are described in the following, and their shortcomings pointed out. The methods are then tried out on a "guinea pig" characteristic and the results critically compared with respect to accuracy, amount of labor, required mathematical tools, and other factors. It is found that no "best" method exists and that all methods require simplifying assumptions and approximations that frequently lead the converter designer to consider practical measurements, or special methods, such as the one provided by the frequency conversion diagram technique.

THE conventional frequency converter has two essential parts; the non-linear element and the local oscillator. The action of the non-linear element or mixer is to combine the incoming wave e_A of angular velocity A with the local oscillator wave e_B of angular velocity B so that one or more product terms result. Each product term yields a sum and a difference component, and the output circuit is arranged to emphasize the current, voltage, or power of a selected sum or difference component, the angular velocity C of which is determined from the relation $C = |pA \pm qB|$, where p and q are any integers. The desired output current is in the following indicated i_C .

The mixer may be linear when operated with zero oscillator voltage, but if so is made non-linear by the application of the oscillator voltage. A rigorous and general mathematical treatment of the non-linear device is more difficult to carry out and apply when the operation is "Class B" or "Class C," and when the frequency is very high. To make a mathematical treatment possible a number of factors, which control the behavior of the non-linear element, must be left out of consideration. Accordingly the results of the mathematical investigation will not tell the full story, and the interpretation of the results must be done with utmost care.

FUNDAMENTAL CONSIDERATIONS

The following treatment is built upon the assumption that the non-linear element of

interest is fully described by a single current-voltage or i - e characteristic. This i - e characteristic may be a tube characteristic, issued by the tube manufacturer, or a curve obtained from measurements on an available physical mixer. In both cases there will be certain errors in the current and voltage coordinates, and these errors are magnified in the readings taken from the curve. Fundamentally there are a number of mathematical methods that can be applied to find the resulting output from a given input. Some of these methods lose their meaning, however, if the primary information is too inaccurate. One way of getting around this difficulty is to assume that the plotted i - e characteristic is exact, i.e., entirely free from error. Several limiting curves representing various amounts of error can then be plotted, and the calculation procedure repeated for each one of them. A comparison of the results gives information about the intervals of variations that can be expected in particular variables, for example the maximum and minimum value of conversion transconductance.

One fundamental idea in the treatment of non-linear devices is to investigate the given, plotted characteristic for the purpose of obtaining its equation in the explicit form $i = f(e)$. This idea seems to leave out of consideration "Class B" and "Class C" operation, as $i = f(e)$ is an analytical function. If, however, the device is provided with a small artificial current in the non-conducting interval of the variational voltage, the non-analytical function of a "Class B" or "Class C" mixer can be transformed into an

analytical one. Such a transformation is actually of practical importance as will be shown later.

Another way of obtaining the desired output component is to make a conventional Fourier analysis of the output current wave. This method has several limitations that will be pointed out in the text. Still another method involves that the "mixer-local oscillator" combination is treated as a unit, the variational transconductance of the unit varying periodically with the oscillator frequency. When encountering this periodic transconductance the incoming voltage yields the well-known amplitude-modulation pattern in the output, indicating the generation of one or more product terms. This method is of considerable practical value.

The following methods are centered on the non-linear element itself, and apply to the types of single input low-ohmic mixers, that have recently become of importance in the very-high frequency and ultra-high-frequency field. The mixer may be considered as operating with short-circuited output, but the treatment applies as well to mixers which have a load impedance of the same value for all frequencies. The given i_e -characteristic is then with advantage transformed into a dynamic i_e -characteristic. The treatment given is not extended to mixer circuits with load impedances of general type, but it is well to remember that in the conventional circuit, with the load impedance existing at one frequency only, the voltage drop across this impedance may be treated as part of the applied voltage.^{1,2} Another case that permits a simplified treatment is the high impedance mixer, as here the load impedance may be neglected in series with the comparatively large average value of the plate resistance.

CURVE-FITTING METHODS

Some useful methods of determining the equation $i=f(e)$ for a given i_e -characteristic are described below. The final form of this equation depends very much upon the method employed. If this form is not explicit in the desired current component, there remains the procedure of

¹ E. G. James and J. E. Houldin, "Diode frequency changers," *Wireless Engineer* 20, 15 (1943).

² E. W. Herold and L. Malter, "Some aspects of radio reception at ultra-high frequency," *Proc. I.R.E.* 31 (August, September, and October, 1943).

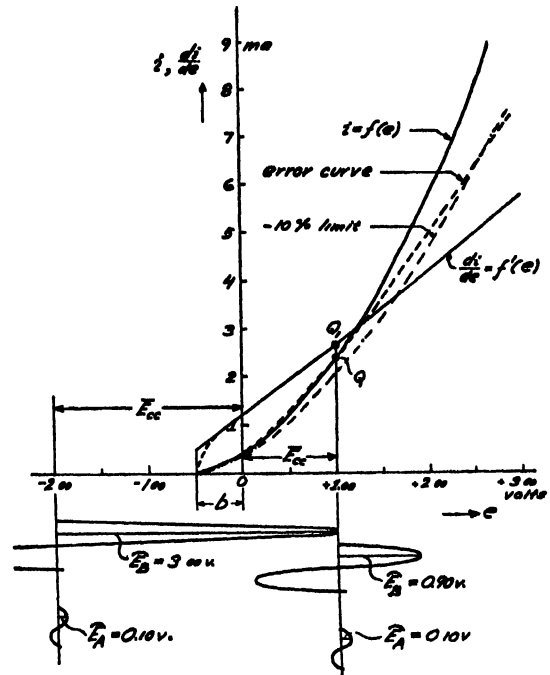


FIG. 1. The "measured" characteristic and its first derivative, illustrating operation in "Class A" and "Class C." The error curve refers to the "log-log straight line" method.

developing such an explicit expression

$$i_C \equiv F(E_A, E_B, A, B).$$

This procedure may involve as much labor as the "curve-fitting" procedure itself.

Before going into a discussion of methods we will study an assumed i_e -characteristic to be used later as a "guinea pig." This assumed characteristic is shown in Fig. 1 and represents a somewhat idealized diode tube. Actually we know the equation for this characteristic, which is, in milliamperes and volts,

$$i = c_1(e+b) + c_2(e+b)^2 + c_3(e+b)^3, \quad (1)$$

$$= 0.50(e+0.50) + 0.70(e+0.50)^2 + 0.01(e+0.50)^3. \quad (2)$$

The variational voltage applied around the Q-point is

$$\Delta e = e_A + e_B = E_{A\max} \cos At + E_{B\max} \cos Bt, \quad (3)$$

where the maximum values are 0.10, respectively, 0.90 volts, the angular velocities $A = 2\pi \cdot 30 \cdot 10^6$ and $B = 2\pi \cdot 27 \cdot 10^6$, and the operation "Class A" with the Q-point coordinate $E_Q = +1.00$ volt. The fundamental operation output will then be

of frequency 3×10^6 c.p.s. Expanding the parenthesis $(\bar{E}_{cc} + \Delta e + b)$ by means of the binomial theorem and collecting proper terms, we obtain the final result

$$i_C = E_{A\max} E_{B\max} [c_2 + 3c_3(\bar{E}_{cc} + b)] \times \cos(A - B)t, \quad (4)$$

which yields the output current 0.0670 ma peak value. (Formula (4) may be directly obtained from the expansion in the appendix with $k_2 = c_2$, $k_3 = c_3$, and $k_4 = 0$.)

The conversion transconductance is defined as

$$g_c = \frac{I_{C\max}}{E_{A\max}} \bigg|_{\text{small values}} \quad E_{C\max} = 0, \quad (5)$$

(in absolute current and voltage values) and becomes 670 μ mhos. The obtained numerical values are in the following treated as being exact so that a comparison can be made between the numerical results obtained with different methods. The methods to be presented follow.

a. $i = f(e)$ Obtained on the Assumption that the Given Characteristic Follows a Square-Law, a $\frac{3}{2}$ -Power Law or Similar Law, Making Possible a log-log Straight-Line Plot.

This method is mainly of interest for ie -characteristics without sharp bends, utilized in "Class A" operation. Assuming an ie -characteristic that extends the amount b on the negative voltage side of the origin, we may write its unknown equation in the form

$$i = f(e) = 0.5K(e + b)^a. \quad (6)$$

In a log-log plot this equation has the form

$$\log i = a \log(e + b) + \log 0.5K. \quad (7)$$

Reproducing the given curve in a log-log plot we may obtain an almost straight line. A substituted straight line has the intercept in the origin $\log 0.5K$ and the slope

$$a = \frac{d[\log i]}{d[\log(e + b)]} = \frac{\log i_1 - \log i_2}{\log(e + b)_1 - \log(e + b)_2}, \quad (8)$$

where subscripts 1 and 2 indicate points on the straight line. Thus the numerical values of K

* For $a = 2$ this becomes the important case of a parabola and the factor 0.5 then makes the constants identical with specific derivatives.

and a can be obtained and the equation $i = f(e)$ determined. The method becomes impractical in cases where $i = f(e)$ cannot be expressed by the simple form (6).

Proceeding from (6) to determine the desired output current i_C for fundamental operation we will avoid the simple case of a equal to 2 and treat the more general case of a unequal to 2. Selecting a Q -point with the positive coordinate $\bar{E}_{cc}$ we obtain for a small voltage variation Δe around the Q -point the variational current

$$\Delta i = -i_Q + 0.5K(\bar{E}_{cc} + b + \Delta e)^a. \quad (9)$$

The parenthesis is then expanded by means of the binomial theorem for the variational voltage Δe given by (3), and all terms containing the product $\cos At \cos Bt$ collected. The result of these operations becomes

$$i_C = 0.25Ka(a-1)E_{A\max}E_{B\max}[(\bar{E}_{cc} + b)^{a-2} + 0.125(a-2) \cdot (a-3)(\bar{E}_{cc} + b)^{a-4} \times (E_{A\max}^2 + E_{B\max}^2) + \dots] \cos(A - B)t. \quad (10)$$

The output current for harmonic operation is obtained in a similar way by collecting terms for the proper higher-order product term. The more terms we include in the expansion the less become the amplitudes, which indicates as a first approximation that harmonic operation is less efficient than fundamental operation.

Applying method (a) to our practical example, we reproduce the given ie -characteristic in Fig. 1 on log-log paper and obtain after rough calculation $K = 2.5$ and $a = 1.7$.⁴ Using Eq. (10) we find the desired output current to be $I_{C\max} \doteq 0.060$ ma, and the error to be -10 percent. Repetition of the procedure will reduce the error to the more permissible value of a few percent.

We may use the same "guinea-pig" characteristic to investigate "Class C" operation, but in this case there is no equation available, the result of which may be used as the exact answer for determining the error in various graphical methods. For "Class C" operation we may locate the Q -point at the coordinate $\bar{E}_{cc} = -2.00$ volts and use an oscillator amplitude of $E_{B\max} = 3.00$ volts. It is understandable that method (a) is not suitable for "Class C" operation, for even if

⁴ Maximum values are in the figures indicated by means of an eyebrow above the symbol.

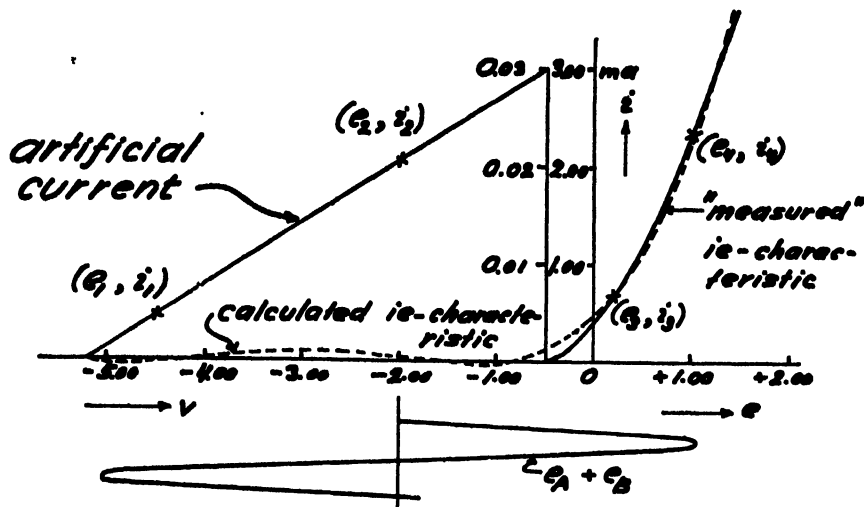


FIG. 2. Method of extending the i_e -characteristic to make possible the application of the variable-coefficient method with four points for "Class C" operation.

we extend the characteristic into the non-conducting interval to obtain an analytical function, it would be most unlikely that the resulting curve should be of such form that it with advantage could be represented by the simple formula (6).

b. $i=f(e)$ Written as a Power Series, the Constants of Which are Determined by the Methods of Variational Coefficients.

We start out with the unknown equation, introducing for simplicity the symbol v to describe the variational coordinate measured from the cut-off point. Then in series representation

$$i=f(v)=\sum_{n=1}^{\infty} k_n v^n, \quad (11)$$

or

$$i=k_1 v+k_2 v^2+k_3 v^3+\cdots, \quad (12)$$

where $k_1, k_2, k_3, \dots$ are the coefficients to be determined by means of the variational coefficient method. The power series method is much more general than the previous method and also handles characteristics with sharp bends. The procedure is to select on the given i_e -characteristic a number of points, which, particularly in case of a sharp-bend curve, have to be carefully grouped—a matter of experience. An equation system is then written down with as many equations, and as many terms in each question, as there are points on the i_e -characteristic. This provides a linear equation system in k , the i - and e -values being read off from the given curve. The

solution of the equation system is generally simple, but may require accurate computation if small differences between large quantities occur. It is recommended that the calculated curve is plotted on top of the given curve so that unpermissible wiggles can be checked. The accuracy of the final result does not necessarily increase with an increased number of terms, as the equation will simultaneously become of higher order and therefore somewhat more difficult to handle, particularly with reference to small differences. It must also be realized that a vaguely known curve does not justify a large number of terms unless the previously mentioned stunt of assuming the curve free from error is made use of.

This method may also be used for the treatment of "Class C" operated devices, if a non-vanishing current is assumed in the non-conducting region of the variational voltage. We will study such a case with reference to the given or "measured" i_e -characteristic reproduced to the right in Fig. 2. The Q-point is located at -2.00 volts. The sum-voltage e_A+e_B , with $E_{A\max}=0.10$ volts and $E_{B\max}=3.00$ volts, is applied around the Q-point. Choosing arbitrarily 1 percent of the maximum current as artificial current value in the cut-off point, we arrive with an assumed straight-line current distribution to the result shown, one hundred times magnified, in Fig. 2. Choosing more or less arbitrarily four points along the total i_e -characteristics, and counting v from the new cut-off point, we can

write down the equation system,

$$\begin{aligned} i_1 &= k_1 v_1 + k_2 v_1^2 + k_3 v_1^3 + k_4 v_1^4, \\ i_2 &= k_1 v_2 + k_2 v_2^2 + k_3 v_2^3 + k_4 v_2^4, \\ i_3 &= k_1 v_3 + k_2 v_3^2 + k_3 v_3^3 + k_4 v_3^4, \\ i_4 &= k_1 v_4 + k_2 v_4^2 + k_3 v_4^3 + k_4 v_4^4. \end{aligned} \quad (13)$$

From the graph we read off the ei -values $(-4.50, 0.00447)$, $(-2.00, 0.02040)$, $(+0.20, 0.7000)$, and $(+1.00, 2.36000)$, which are given with an accuracy justified by the assumption that the given curve is exact. As the equation system is linear in k we obtain after straightforward computation work the results $k_1 = -0.130$, $k_2 = +0.244$, $k_3 = -0.101$, and $k_4 = +0.012$. For checking purposes the resulting curve has been roughly plotted in Fig. 2, and it is seen that the agreement between the given and the computed characteristic is quite good. Thus we can accept as a working empirical equation

$$i = -0.130v + 0.244v^2 - 0.101v^3 + 0.012v^4. \quad (14)$$

Introducing here

$$v = V + E_{A\max} \cos At + E_{B\max} \cos Bt, \quad (15)$$

where $V = 3.20$ volts, expanding the terms in (14) and collecting and grouping terms we obtain (see Appendix with $(E_{oc} + b) = V$)

$$\begin{aligned} i_C &= E_{A\max} E_{B\max} [k_2 + 3k_3 V \\ &+ 6k_4 (V^2 + \frac{1}{2} E_{A\max}^2 + \frac{1}{2} E_{B\max}^2)] \cos (A - B)t, \end{aligned} \quad (16)$$

which yields the numerical value $I_{C\max} \doteq 0.052$ ma. This is the intermediate frequency output for fundamental operation, "Class C." The conversion transconductance is approximately 520 μ mhos.

To provide figures for the table in the latter part of the discussion calculations were as well carried out for the "Class A" case. The result is $I_{C\max} = 0.0663$ ma, the error being -1 percent.

§. *$i = f(e)$ Written as a Power Series, the Constants of Which are Determined by the Method of Taylor's Series*

In this method we determine the coefficients as derivatives in the Q -point, and it follows that if there is a sharp bend in this point or anywhere else in the interval so that the function becomes non-analytical, the derivative cannot be deter-

mined and the method breaks down. Thus the method applies to smooth curves only, and is restricted to "Class A" operation, as an artificially extended ie -characteristic for "Class C" operation scarcely fulfills the requirements on a smooth curve. (The extended curve in Fig. 2 would yield second and higher-order derivatives of zero value.)

Generally we may write Taylor series for a single input device

$$\Delta i = \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{\partial}{\partial e} \right)^n i, \quad (17)$$

where $\partial/\partial e$ operates upon i . For two variables Δe_1 and Δe_2 , so applied that

$$\Delta e = \Delta e_1 + \Delta e_2, \quad (18)$$

we obtain

$$\begin{aligned} \Delta i &= a \Delta e_1 + b \Delta e_2 + c \Delta e_1^2 + d \Delta e_1 \Delta e_2 \\ &+ f \Delta e_2^2 + g \Delta e_1^3 + \dots \end{aligned} \quad (19)$$

The coefficients in this series have the values

$$\begin{aligned} a &= \frac{\partial i}{\partial e_1}, & c &= \frac{\partial^2 i}{\partial e_1 \partial e_2}, \\ b &= \frac{\partial i}{\partial e_2}, & f &= \frac{1}{2} \frac{\partial^2 i}{\partial e_2^2}, \\ c &= \frac{1}{2} \frac{\partial^2 i}{\partial e_1^2}, & g &= \frac{1}{6} \frac{\partial^3 i}{\partial e_1^3}, \\ &\dots & & \dots \end{aligned} \quad (20)$$

where several numerical identities occur as we deal with one current-voltage function only. (Partial derivatives are used unconventionally to indicate that the evaluation is made in the Q -point). For fundamental operation and with the variational voltage (18) expressed by (3) the useful i - f output becomes

$$\Delta i = \dots + \frac{\partial^2 i}{\partial e_1 \partial e_2} E_{A\max} \cos At \cdot E_{B\max} \cos Bt, \quad (21)$$

so that

$$i_C = \frac{1}{2} \frac{\partial^2 i}{\partial e_1 \partial e_2} E_{A\max} E_{B\max} \cos (A - B)t. \quad (22)$$

By plotting the slope and measuring the slope of the slope of the given ie -characteristic in Fig. 2 we obtain $\partial^2 i / \partial e_1 \partial e_2 = 1.50$ ma/volts², so that with $E_{A\max} = 0.10$ volt and $E_{B\max} = 0.90$ volt the result becomes $I_{C\max} = 0.0675$ ma. The

error in this result is approximately +1 percent, which is better than what can generally be expected from a slope-of-the-slope method (the error may easily have amounted to 5 or 10 percent).

In more complicated applications the following treatment may prove useful. We will write the total instantaneous current, with $\partial/\partial e$ operating upon the current i ,

$$i = f(e) = f(\bar{E}_{cc}) + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{\partial}{\partial e} \right)^n i. \quad (23)$$

Here $f(\bar{E}_{cc})$ is the direct-current component produced in the Q -point by the bias voltage $\bar{E}_{cc}$. Rather than to expand (23) we will make use of the following transformation

$$f(e) = f(\bar{E}_{cc} + \Delta e) = e^{(\partial/\partial e)\Delta e} f(\bar{E}_{cc}), \quad (24)$$

which can be proven by series expansion. We obtain for $\Delta e = \Delta e_1 + \Delta e_2$,

$$i = e^{(\partial/\partial e_1)\Delta e_1} \cdot e^{(\partial/\partial e_2)\Delta e_2} \cdot f(\bar{E}_{cc}). \quad (25)$$

Thus the current is generally obtained as the product of two factors, each one with an exponent which is a specified derivative of the given current-voltage function evaluated in the Q -point. As Δe_1 and Δe_2 are both cosinusoidal, the e -functions become periodical and can be expanded in Fourier series with Bessel function coefficients. These steps will not be carried through as the following method involves practically the same developments (see Eqs. (27) and (28)).⁵

d. Empirical Equation Written as an ϵ -Series and Developed in Fourier Series with Bessel-Function Coefficients

As an alternative to the previously given power series, we may use the ϵ -series

$$i = \sum_{n=1}^{\infty} A_n \epsilon^{a_n v} = A_1 \epsilon^{a_1 v} + A_2 \epsilon^{a_2 v} + \dots, \quad (26)$$

where, as before, the coordinate is counted from the cut-off point. For a smooth "Class A" characteristic of roughly exponential form this series can with advantage be utilized as equation.

⁵ A. C. Bartlett, "Calculation of modulation products," *Phil. Mag.* 16, 843 (1933) and 17, 628 (1934).

Two terms involve four coefficients to be determined. Again "Class C" operation would necessitate artificial extension of the given i -characteristic, only feasible in the few cases where the resulting characteristic remains essentially exponential.

It is desirable that the series in (26) be transformed into a new series, in which the product terms appear in more obvious form, so that the sum and difference frequency components can be quickly identified. The final series will be explicit in the desired current component only if the applied voltages are independent of current. The variational voltages in a modern u.h.f. diode converter circuit sometimes depend upon various current components in such a way that the problem becomes complicated beyond engineering consideration.^{1,2}

If the applied voltage is of form (15) we may write down a general term m in expansion (26)

$$A_m \epsilon^{a_m v} = A_m \epsilon^{a_m V} \cdot \epsilon^{a_m E_{A\max} \cos At} \cdot \epsilon^{a_m E_{B\max} \cos Bt}. \quad (27)$$

To expand a factor of the form $\epsilon^{k \cos \omega t}$ we may make direct use of the relation

$$\begin{aligned} \epsilon^{k \cos \omega t} &= J_0\left(\frac{k}{j}\right) \\ &- 2\left[J_2\left(\frac{k}{j}\right) \cos 2\omega t - J_4\left(\frac{k}{j}\right) \cos 4\omega t + \dots\right] \\ &+ 2j\left[J_1\left(\frac{k}{j}\right) \cos \omega t - J_3\left(\frac{k}{j}\right) \cos 3\omega t + \dots\right], \end{aligned} \quad (28)$$

where $J_0(k/j)$ is the Bessel function of the first kind and 0th order and $J_\nu(k/j)$ the Bessel function of the first kind and ν th order, k/j , or rather jk with proper sign, being the argument. Using mathematical texts and tables for the Bessel functions, we proceed to express each one of the last two factors in (27) by means of (28) with $m=1$, $m=2$, etc. Carrying out these steps and grouping and collecting terms for fundamental operation, we obtain

$$i_C = I_{C\max} \cos(A-B)t, \quad (29)$$

where

$$\begin{aligned} I_{C\max} &= 2[A_1 \epsilon^{a_1 V} j J_1(j a_1 E_{A\max}) j J_1(j a_1 E_{B\max}) \\ &+ A_2 \epsilon^{a_2 V} j J_1(j a_2 E_{A\max}) j J_1(j a_2 E_{B\max})]. \end{aligned} \quad (30)$$

Applying this method to our practical example

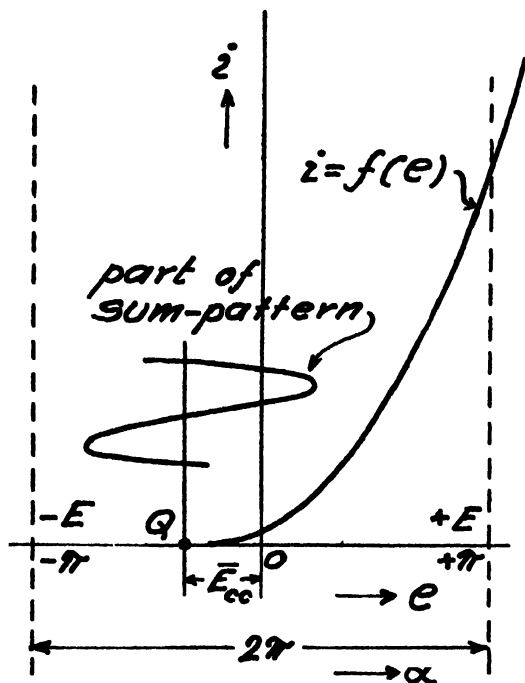


FIG. 3. Obtaining an empirical equation for a measured ie -characteristic by means of single period Fourier analysis in terms of voltage.

we will find that if all factors are known the procedure mainly simplifies to the use of tables of Bessel functions. We may, however, run into considerable difficulties when trying to determine the coefficients $A_1, A_2, \dots, a_1, a_2, \dots$. Limiting ourselves to two terms in the practical example we will here employ a graphical method for determining the four coefficients by replotting the given ie -characteristic with logarithmic current axis. If this construction is carried out it is found that the resulting curve bends downwards so that if we approximate its initial part with a straight line of the form

$$\log i_1 = 0.4343a_1v + \log A_1, \quad (31)$$

we shall have to subtract the ordinates for a second similar line, covering the upper part of the given characteristic. Carrying out the graphical work and reading off the constants for the two straight lines we obtain for the given practical application the approximate values

$$A_1 \pm 0.40, \quad a_1 \pm 1.35, \quad A_2 \pm -0.04, \quad \text{and} \quad a_2 \pm 1.97.$$

Using formulas (29) and (30) we obtain $I_{Cmax} = 0.05$ ma. The error in the answer is -25 per-

cent and is representative for the inaccuracy that can be expected in applications where the given ie -characteristic is not essentially of exponential form.

e. $i = f(e)$ Obtained by Means of the Trigonometric Polynomial Method

Let us consider the given characteristic in Fig. 3 for the general case of "Class B" or "Class C" operation. Applying the sum-pattern $e_A + e_B$ around a chosen Q -point we may select a voltage interval $-E$ to $+E$ and treat this interval as a single period of a repeated wave.⁶ By means of "single period" Fourier treatment we can then determine the equation for the extended ie -characteristic, or rather the path of operation between $-E$ and $+E$, and thus obtain, with the voltage coordinate e expressed by $\alpha = (\pi/E)e$,

$$i = f(\alpha) = b_0 + \sum_{n=1}^{\infty} (b_n \cos n\alpha + a_n \sin n\alpha). \quad (32)$$

Here

$$b_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(\alpha) \cos n\alpha d\alpha, \quad (33)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(\alpha) \sin n\alpha d\alpha. \quad (34)$$

Making use of the relation $\alpha = (\pi/E)e$, where

$$e = \bar{E}_{cc} + E_{Amax} \cos At + E_{Bmax} \cos Bt, \quad (35)$$

(with $\bar{E}_{cc}$ indicating the bias voltage), we may solve for i in (32) and write out a series expansion with terms of type $\cos$ (amplitude $\cdot \cos$ argument) or $\sin$ (amplitude $\cdot \sin$ argument), so that we with advantages can use Bessel functions for the amplitudes. Thus the desired product term can be quickly determined by means of a table of Bessel functions, and the work involved is actually less than what would at first be surmised from (32).

The reason why a comparatively large number of terms must be employed is the sudden jumps in the curve, causing oscillations in the empirical curve. Professor W. L. Barrow has shown a method⁷ by means of which the number of terms

⁶ A. Preisman, *Graphical Constructions for Vacuum Tube Circuits* (McGraw-Hill Book Company, Inc., New York, 1943).

⁷ W. L. Barrow, "Contribution to the theory of non-linear circuits with large applied voltages," *Proc. I.R.E.* 22, 981 (August, 1943).

can be reduced. He extends the *ie*-characteristic to become part of an artificial, smooth characteristic that approaches a sine wave or cosine wave in shape, and then utilizes only the part of the curve covered by the variational voltage. No computation work with reference to the trigonometric polynomial method has been carried out, as reference (7) contains extensive calculations that closely resemble the ones we should have to go through in such a treatment.

OTHER METHODS

f. Conventional Fourier Analysis of the Total Output Current, the Function $i=f(e)$ Remaining Unknown

If we project the sum-pattern $E_{A\max} \cos At + E_{B\max} \cos Bt$ on a given *ie*-characteristic, we obtain a distorted sum-pattern, such as the one shown to the right in Fig. 4. It is assumed that A and B are related in such a way that $(A-B)$, A , B , and $(A+B)$ are all component values in one and the same harmonic series, $(A-B)$ being the fundamental. Only in special cases would practical A and B values fulfill such a requirement. This means that $(A-B)$ is generally not the fundamental, so that Fourier analysis performed with $(A-B)$ as fundamental becomes erroneous. The true fundamental is in the general case some component $|pA \pm qB|$ with particular numerical values for the integers p and q . As an example assume that A and B are related by an irrational number such as $\sqrt{2}$. Performing the stunt of approximating $\sqrt{2}$ to 1.4 we may transform the problem and write $10A = 14B$, meaning that the cycle will close after 14 cycles of A (or 10 cycles of B), so that a repeating wave is obtained that can be analyzed in conventional way. Such a transformation may, however, increase the amount of labor beyond engineering considerations. In other applications we may be able to change the numerical value of A or B until a repetition cycle of small enough period is produced.

When the method applies, the output current is of the form

$$i(t) = b_0 + \sum_{n=1}^{\infty} (b_n \cos n\omega t + a_n \sin n\omega t), \quad (36)$$

where a coefficient such as b_n is conventionally

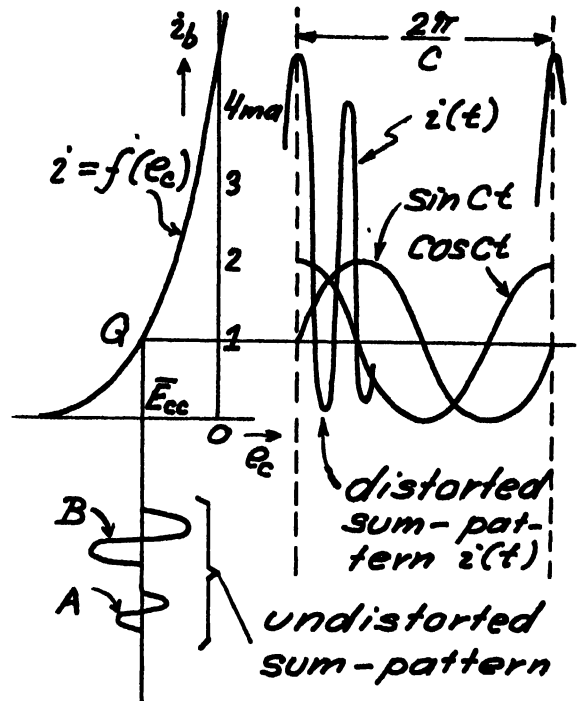


FIG. 4. Sum-pattern $e_A + e_B$ impressed on a measured *ie*-characteristic with the resulting current variation shown to the right. This is a special case, suitable for Fourier analysis.

written

$$b_n = \frac{2}{T} \int_{-T/2}^{+T/2} i(t) \cos n\omega t dt. \quad (37)$$

To determine such a component we may, for example, perform a graphical integration. For the simple case of fundamental operation with $A-B=C$, (37) takes the form

$$b_1 = \frac{2}{T} \int_{-T/2}^{+T/2} i(t) \cos Ctdt. \quad (38)$$

Superimposing a unit amplitude cosine wave, and employing a planimeter or counting squares, we obtain the integral above in agreement with Fig. 4, and thus find $b_1 = 2y$, where y is the height of a rectangle of the same area as is indicated by the planimeter. The coefficient a_1 is obtained in a similar way (or is conveniently arranged to have zero value), whereafter $I_{C\max} = (b_1^2 + a_1^2)^{1/2}$.

Theoretically at least the Fourier analysis can be applied directly to the *ie*-characteristic without the necessity of plotting the output wave. This harmonic analysis procedure, so practical when applied to sine-wave driven

power amplifier tubes,^{8,9,10} seems not very attractive for sum-pattern operated converters.

Using conventional Fourier analysis in the practical example, "Class A" operation, we obtain $I_{C\max} = 0.09$ ma. The error in this value is very high, +35 percent, because of the approximative nature of the graphical integration. An accurate value is hard to obtain in this example, as we deal with a small difference between large plus and minus areas.

g. Treatment When the A-Wave Only is Applied to the Non-Linear Element, the ie-Characteristic of Which is Replaced by a B-Wave Controlled Instantaneous Characteristic

A frequency converter may be defined as a device with periodically varying transadmittance, the periodicity of the variation determining the frequency of each possible output component. Neglecting the out-of-phase component we may develop the transadmittance as the transconductance $g_m(t)$ in a Fourier series with the fundamental frequency $B/2\pi$, the local oscillator voltage being

$$e_B = E_{B\max} \cos Bt. \quad (39)$$

Thus

$$g_m(t) = b_0 + \sum_{n=1}^{\infty} (b_n \cos nBt + a_n \sin nBt). \quad (40)$$

The transconductance $g_m(t)$ will contain harmonics even if the applied oscillator voltage, as in (39), is free from harmonics. The coefficients in this series are defined in conventional manner.

Applying this method to our practical example, we obtain for "Class A" operation, with the Q-point at plus one volt and $E_{B\max} = 0.90$ volt, a slope $(g_{m1})_{\max} = 1.33$ ma/v. As the slope curve happens to be almost a straight line, there will be a negligible amount of harmonics in the variational conductance. Thus, with $b_0 = g_{m0}$ in (40),

$$g_m(t) = g_{m0} + (g_{m1})_{\max} \cos Bt. \quad (41)$$

Fundamentally the resulting variational current is obtained as the product of the variational

conductance and the applied variational voltage, so that, leaving any existing bias voltage out of consideration, we obtain

$$\Delta i = g_m(t) \cdot \Delta e_A \\ = (g_{m0} + (g_{m1})_{\max} \cos Bt) E_{A\max} \cos At, \quad (42)$$

which yields the difference frequency term

$$i_C = \frac{1}{2} (g_{m1})_{\max} E_{A\max} \cos (A - B)t. \quad (43)$$

With an A-wave amplitude of 0.10 volt, formula (43) yields $I_{C\max} = 0.0665$ ma. The error here is less than -1 percent, which is a permissible value.

"Class C" operation requires a more elaborate analysis. The slope curve is now discontinuous and again we employ Fourier analysis to find the desired conductance harmonic (counting the fundamental as the first harmonic). The graphical integration method described under Section (f) may be used, or we may employ conventional schedule analysis. Using the Chaffee harmonic analysis 9-point method^{8,10} we obtain for fundamental operation

$$(g_{m1})_{\max} = \frac{1}{8} [S_1 + 3S_2 + S_4], \quad (44)$$

where the ordinates S_1 , S_2 , and S_4 have the respective values 2.12, 2.67, and 0 ma. Using formulas (42) and (43) we then obtain $I_{C\max} = 0.056$ ma. (The graphical integration method gave the answer 0.057 ma, and proved to be the slower method of the two.)

CONCLUSIONS

There are other methods and combination of methods that could be considered but the presented material is sufficient to shed light on the questions of interest. Table I gives the com-

TABLE I.

Method	$I_{C\max}$ in ma	Error in percent
a. log-log straight-line plot	0.060	-10
b. variational coefficients	0.0663	-1
c. Taylor series coefficients	0.0675	+1
d. s-series, Bessel coefficients	0.05	-25
f. Fourier analysis of output	0.09	+35
g. variational conductance	0.0665	-1

putation results obtained with the various methods.

The error values have been calculated with

⁸ E. L. Chaffee, *Theory of Thermionic Vacuum Tubes* (McGraw-Hill Book Company, Inc., New York, 1933).

⁹ E. L. Chaffee, "The operating characteristics of power tubes," *J. App. Phys.* 9, 471 (June, 1938).

¹⁰ E. L. Chaffee, "Power tube characteristics," *Electronics* 11, No. 6 (June, 1938).

0.0670 ma as the correct value, obtained for the test or "guinea pig" characteristic from Eq. (2).

When comparing the results, the writer concludes that the natural form of the unknown mathematical expression for the studied characteristic and the actual shape of this characteristic favors certain methods and handicaps others. To make this point more clear, suppose we started out to investigate a characteristic that happens to be fairly exponential in its form. Two ϵ -terms would then describe this characteristic with good accuracy, and method *d* gives an error of only a few percent, while some of the other methods with tabulated errors of the order of one percent may give very large errors. As another example, assume that the form of the characteristic, in amplitude and frequency values, was very suitable for investigation by means of Fourier analysis. Then method *f* may give an error of only a few percent, while (if the input voltage effects the transconductance variation) a method like *g* may give a large error. More enlightening examples could be brought up if we extended the discussion to the treatment of discontinuities, such as sharp bends.

The judgment of the proper method becomes still more difficult if the discussion is extended to harmonic operation and to problems concerning loading conditions (to the extent we can cope with such problems). Further, there is no sharp definition of what is meant by the "best method," for there are many factors that enter into the comparison, for example accuracy of final result, amount of labor involved, level of required skill, type of required skill, type of computation equipment required, possibilities of extending the method, etc. It follows that a successful comparison of different methods can only be made if the scope of the work, and the nature of the judgment, is clearly specified from the beginning, and in addition the computation work is carried far outside the limits indicated by this paper. A number of "guinea pig" characteristics must be employed and a number of operating conditions must be studied. Actually, the serious investigator must be prepared to have at hand a computation office to take care of the landslide of calculations that result from a general approach to the problem.

The above discussion concerns mainly "Class A" operation. The extension of the treatment to "Class B" and "Class C" operation brings up a number of additional questions. The computation work for one analytical method, the variational coefficient method, and one graphical method, the variational conductance method, has here been carried out. The former gives a fundamental operation output current of 0.052 ma and the latter for the same quantity 0.057 ma. The error in one, considering the other correct, is only 5 percent, which is a good result, considering the straightforward manner of computation. Some of the presented methods may not apply at all to "Class C" problems, while others can be made to apply, for example by means of the mathematical stunt of adding an artificial current in the non-conducting region. (A comparison of "Class A" and "Class C" operation from the computation values in the practical example would not be fair as we are not making the most of the "Class C" operation.)

To summarize the most important points, there does not exist any one particular method to determine *i-f* output current and conversion transconductance, which method is general, quick, accurate and reliable. This is particularly true as most converters operate in "Class B" or "Class C." There are, however, certain methods, such as the variational conductance method, which for many applications yield a fairly accurate answer in a short time, and such a method is therefore used more frequently than other methods. The bottleneck in the use of the various methods is that we in complicated cases are forced to deal with so strongly simplified conditions that the close connection between our computation results and the true characteristics of the physical device are obscured. There are many factors, which have a bearing on the result, and if all these factors were included in our calculations, the problem would become too complicated for a straightforward theoretical treatment.

It is therefore natural that direct measurement methods are sometimes preferred for computation methods, and the experienced converter designer frequently finds himself involved in extensive practical measurements, in which mathematical investigations of the type presented in

this paper are coming in here and there merely as aids in the development work.

It is at this point the Chaffee Conversion Diagram Method calls for attention, as this method may be worth consideration between all-out calculations and all-out measurements. The conversion diagrams encourage systematic investigation and describe important relations under practical operating conditions for the nonlinear element. Thus some of the resistive and reactive properties of the nonlinear element, which are very difficult to take into account in calculations, may be automatically taken into account by a conversion diagram. Useful curves, such as the i - f output power as function of the oscillator voltage, can be directly plotted from the diagrams. The writer will not in this connection carry the discussion of the conversion diagram method further. For an estimation of its value as a tool for the converter designer the reader is referred to the references.¹¹⁻¹⁴

ACKNOWLEDGMENT

This paper provides the introductory part of a thesis for doctor's degree at Harvard University, and the author wishes to thank the Director of Cruft Laboratory, Professor E. L. Chaffee, for having made this work possible.

¹¹ H. Stockman, "A treatment of non-linear devices based upon the theory of related linear functions," J. App. Phys. 14, 645 (1943).

¹² H. Stockman, "Use of frequency conversion diagrams," Proc. I.R.E. 32, 679 (November, 1944).

¹³ H. Stockman, "UHF converter analysis," Electronics 18, 140 (February 1945), and Proc. Nat. Electronics Conference, Chicago, 1, 242 (1944).

¹⁴ H. Stockman, "Frequency conversion diagrams for triode tube mixers," J. App. Phys. 16, 639 (October, 1945).

APPENDIX

Obtain the desired product term from the power series expansion in (12).

The given expansion is of the form

$$i = \sum_{n=1}^{\infty} k_n v^n = k_1 v + k_2 v^2 + k_3 v^3 + k_4 v^4 + \dots, \quad (12)$$

where we will neglect higher order terms. Introducing $v = (e + b)$ we obtain the series

$$i = M_0 + M_1 e + M_2 e^2 + M_3 e^3 + M_4 e^4 + \dots, \quad (12.1)$$

where

$$M_0 = k_1 b + k_2 b^2 + k_3 b^3 + k_4 b^4 + \dots,$$

$$M_1 = k_1 + 2k_2 b + 3k_3 b^2 + 4k_4 b^3 + \dots, \text{ etc.}$$

Introducing

$$e = \bar{E}_{cc} + E_{A\max} \cos At + E_{B\max} \cos Bt \quad (12.2)$$

in (12.1) we obtain, among other terms,

$$i = \dots + E_{A\max} E_{B\max} [M_2 + 3M_3 \bar{E}_{cc} + 6M_4 (\bar{E}_{cc}^2 + \frac{1}{4} E_{A\max}^2 + \frac{1}{4} E_{B\max}^2)] \times \cos (A - B)t, \quad (12.3)$$

so that the desired difference-frequency current becomes, with the M -coefficients expressed in k and b ,

$$i_C = E_{A\max} E_{B\max} [k_2 + 3k_3 (\bar{E}_{cc} + b) + 6k_4 \{ (\bar{E}_{cc} + b)^2 + \frac{1}{4} E_{A\max}^2 + \frac{1}{4} E_{B\max}^2 \}] \times \cos (A - B)t. \quad (16)$$

This final expression is made use of in different ways in the text.

Filmless Sample Mounting for the Electron Microscope

JOHN H. L. WATSON*

Shawinigan Chemicals Limited, Shawinigan Falls, Quebec

(Received November 15, 1945)

A method is described for supporting airborne particles upon thin glass fibers where they build up in chains and may be examined in the electron microscope. With no supporting membrane, the contrast and resolution are increased in the micrographs. The specimens are particularly useful for stereoscopic studies. A special reference is made to the study of carbon particles mounted in this fashion. The carbons examined are formed by pyrolysis of acetylene and occur naturally in pronounced chains. Reproducible particle size data for the carbons can be secured by measurements made on individual particles at $\times 200,000$. Evidence for the crystalline nature of this carbon is rendered visible in the micrographs.

THE thin supporting film upon which electron microscope samples are mounted for examination in the instrument causes some scattering of the electrons. This scattering in the film results in lowered contrast and less than optimum resolution in the image. If it is possible to eliminate the supporting film in some cases and to mount the samples directly in the vacuum, the possible contrast and the practical resolution can be increased.¹ This is especially important when dealing with particularly minute particles of low atomic weight where the amount of scattering in the specimen itself may be of the same order as that in the film. This present discussion is the result of experiments designed to support such samples in free space. The methods have already proved useful in several problems and may suggest applications to others working with electron microscopes.

The first step in the method consists in supporting several fine fibers across the mouth of the usual electron microscope specimen holder. These can be cemented in place with rubber cement or metallic paint, or they can be held mechanically by the pressure of the cap. Fibrous materials which have been found useful for the purpose are glass, rubber, and asbestos. Glass wool fibers of a range of sizes down to one micron diameter are obtainable. Rubber fibers of even smaller diameter can be made by drawing out rubber cement.** The production of tough resinous

fibers of cross-sectional diameter less than one micron may provide a source of the finest fibers usable in the mounting process here outlined.

The holder with the fibers attached is plugged into a rubber tube leading to the source of the fine particles which are to be mounted upon the fibers. The source material, which may be a dry powder, a suspension, or a true solution is atomized to form an aerosol which is drawn past the fiber by a suitable pumping arrangement. As the airborne particles go past, many of them are caught and a deposit, usually in a chain-like form, begins to build up. It is interesting to note that the 'new fibers' composed of the collected particles, themselves, act as very efficient and perhaps more efficient collectors than the original fibers. The obvious tendency of the particles mounted in this way to form chains can be seen

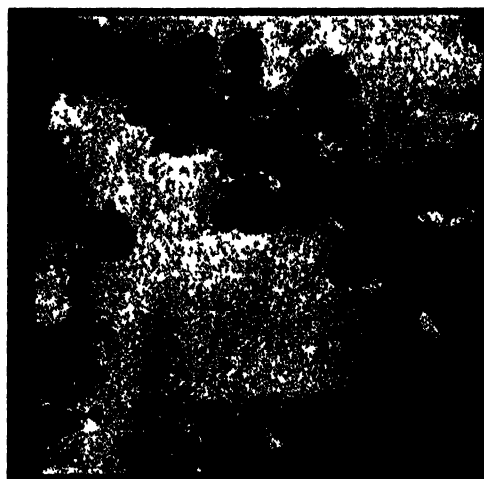


FIG. 1. Particles from a dispersion of methylene blue caught upon the fibers of a thin piece of lens paper. $\times 4,700$ (University of Toronto).

* The method described was noted first when the author was with the electron microscope laboratory of the University of Toronto.

¹ James Hillier, A.S.T.M. Symposium on new methods for particle size determination in the subsieve range, p. 91 (1941).

** Dunlop rubber cement.

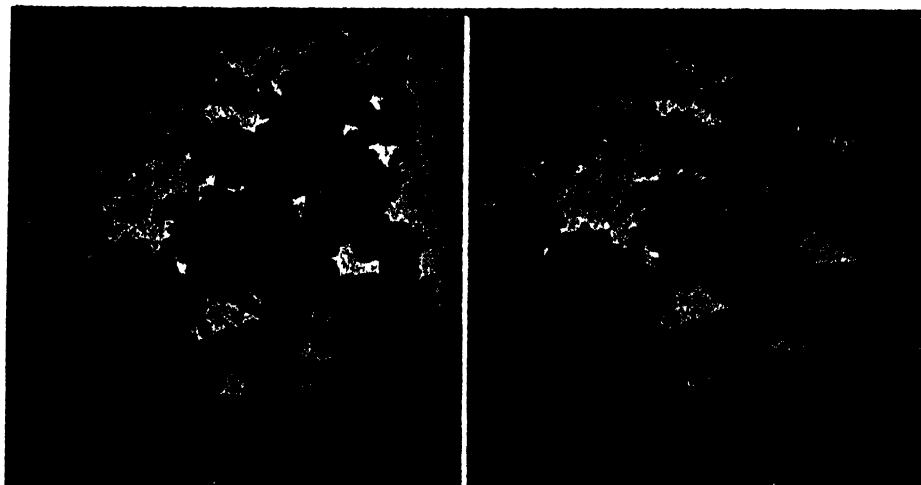


FIG. 2 A stereoscopic pair of a dispersion of Shawinigan carbon black supported in free space $\times 22,000$ (Shawinigan Chemicals Ltd)

in Figs. 1 and 3. The speed past the fiber, the collecting period, and the size of the fiber affect the density of the deposit.

Many variations of this general method can be employed. The author has used thin lens paper (rice paper) for mounting particulate samples by drawing the aerosols through the paper, Fig. 1. The particles caught in the lens paper are of methylene blue which were atomized from a true solution in water. Thin asbestos paper will also serve.

The mounting obtained by the above methods has been found to be remarkably stable during mechanical handling if ordinary care is exercised and may be used for all electron microscope

techniques without harmful effects to the specimen. Electrical charging of a specimen so supported is sometimes a problem. Usually, quite sufficient intensity and contrast for focusing can be achieved without any charge effect being noticed. If charge 'bulges' do appear, or if specimen deterioration does take place, a 200-mesh specimen screen (without a film) put either under or over the fibers before or after the mounting of the sample will dissipate the charge almost completely. It is wise in general to start with low beam and condenser current and to raise each carefully for focusing.

Much can be learned with this method of the actual way in which the particles build up the



FIG. 3. Electron micrograph and diagrammatic representation of sodium chloride crystals caught upon a rubber fiber. $\times 2,300$ (University of Toronto).



FIG. 4. Shawinigan carbon black. $\times 30,000$ (Shawinigan Chemicals Ltd.).

one upon the other. The amount of depth in the object is much greater than that present in a sample mounted upon a film and excellent stereoscopic micrograph pairs can be obtained. Figure 2 is a stereoscopic view of carbon particles supported in free space. The tendency of the material to charge up electrically is useful for studying it three dimensionally, because different sections of the specimen can be made to move relatively to each other by slight variations of intensity. The structure of apparent aggregates can thus be studied visually at first hand on the screen of the microscope.

Figure 3 shows a deposit of small crystals of sodium chloride caught upon a rubber fiber; a diagrammatic representation is also shown of how these crystals when caught build up into a chain.²

THE USE OF THIS METHOD WITH COLLOIDAL CARBON SPECIMENS

Carbon micrographs often lack contrast and sufficient resolution. Indeed, the finer carbon

² E. F. Burton, *Nature* 152, 540 (1943).

blacks are extreme tests for the electron microscope. In order to achieve sufficient image contrast with them, it is often necessary to use an objective diaphragm. No diaphragm was used in obtaining any of the micrographs presented here. The Shawinigan carbon black has a well-known chain structure, a property important for its use in battery manufacture and in the production of thermally and electrically conductive rubbers. This structure is illustrated in Figs. 4 and 5. The chains are stretched out in space from the supporting fiber with contrast and resolution which allow a very accurate measure of particle size distribution to be made, where it might not otherwise be possible to do so. Of course only those particles projecting beyond the edge of the fiber can be counted. A sufficient number of particles can be counted in the chains, and those caught immediately at the fiber can be disregarded. It will be observed that both large and small particles are caught and there is little likelihood of sampling taking place. The plates are projected upon a screen at

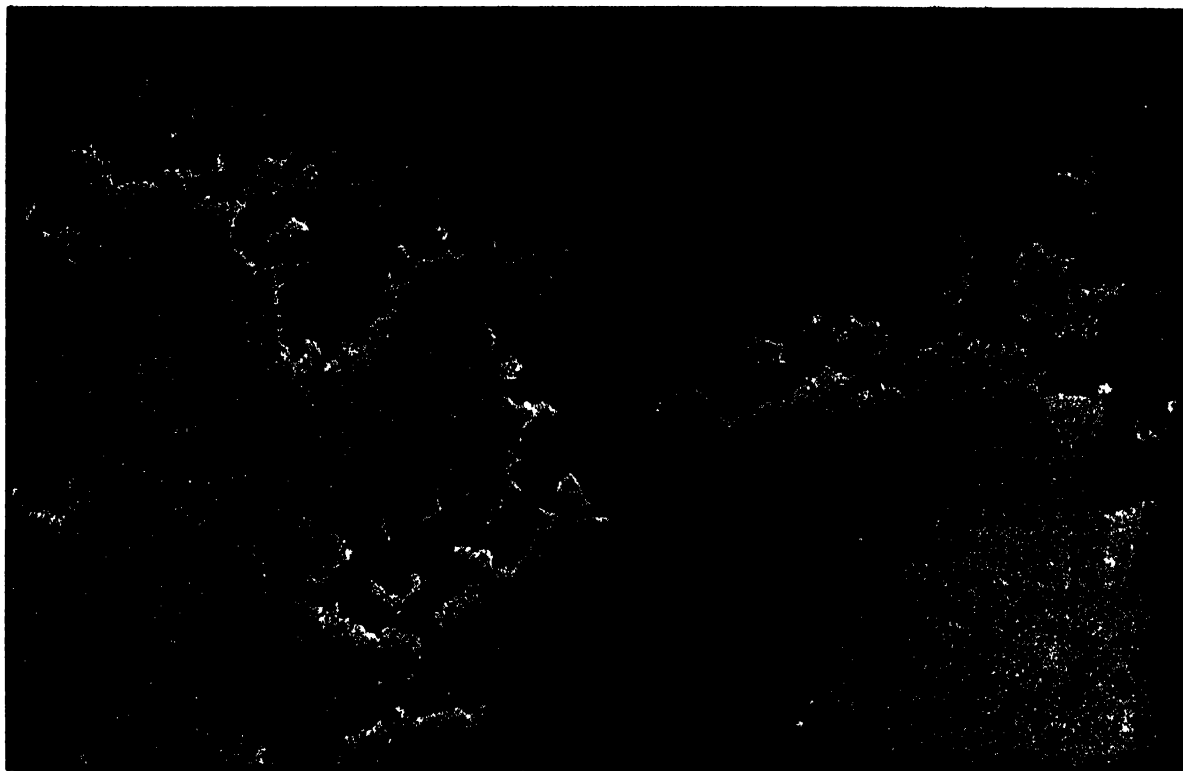


FIG. 5. Shawinigan carbon black. $\times 30,000$ (Shawinigan Chemicals Ltd.).

$\times 200,000$, and the carbon particles are classified, as they are counted, into ranges of 50A. With the electron microscope, it is not feasible to count all the particles along the whole fiber but a number of fields give a satisfactory result. It has been found that ten good exposures taken at random of two preparations of the same sample give reproducible results for that sample.

Figures 4 and 5 are both micrographs of colloidal carbon black produced by the pyrolysis of acetylene. The most definite evidence for the crystalline nature of carbon black has come from x-ray and electron diffraction studies which have suggested that a single acetylene carbon black *particle* consists of a cluster of *pseudo-crystals* (or parallel layer groups) some 40 to 50A in length and 20 to 30A in height, these in turn being composed of thin carbon *platelets* (or ultimate graphite layers) lying parallel to and equidistant from each other but randomly oriented about the perpendicular to their surfaces.³

³ J. N. Mrgudich and R. C. Clock, *Electrochem. Soc.* 29, 329 (1944).

In these electron micrographs the crystalline nature of the carbon is made manifest. At locations *A* there are crystals which appear to be thin platelets, lying in a "flat" position, i.e., with their flat faces approximately perpendicular to the direction of the electron beam. At locations marked *B* the crystals appear to be "edge-on" to the beam. Over the micrographs positions of the crystals intermediate to these two extremes can be noted. Figures 6 and 7 are enlargements of portions of Figs. 4 and 5, respectively. Those few locations marked are representative of many such crystalline occurrences. The high chemical purity of the black leaves little doubt that these crystals are actually carbon.

The edge-on position provides an effectively thicker sample to scatter the electrons, and consequently the apparent "rods" in the micrographs have good contrast. In the flat position however the crystals offer much less scattering power and the contrast is less. From measurements made on the rods the crystal thicknesses are anywhere from 30 to 100A. Therefore the

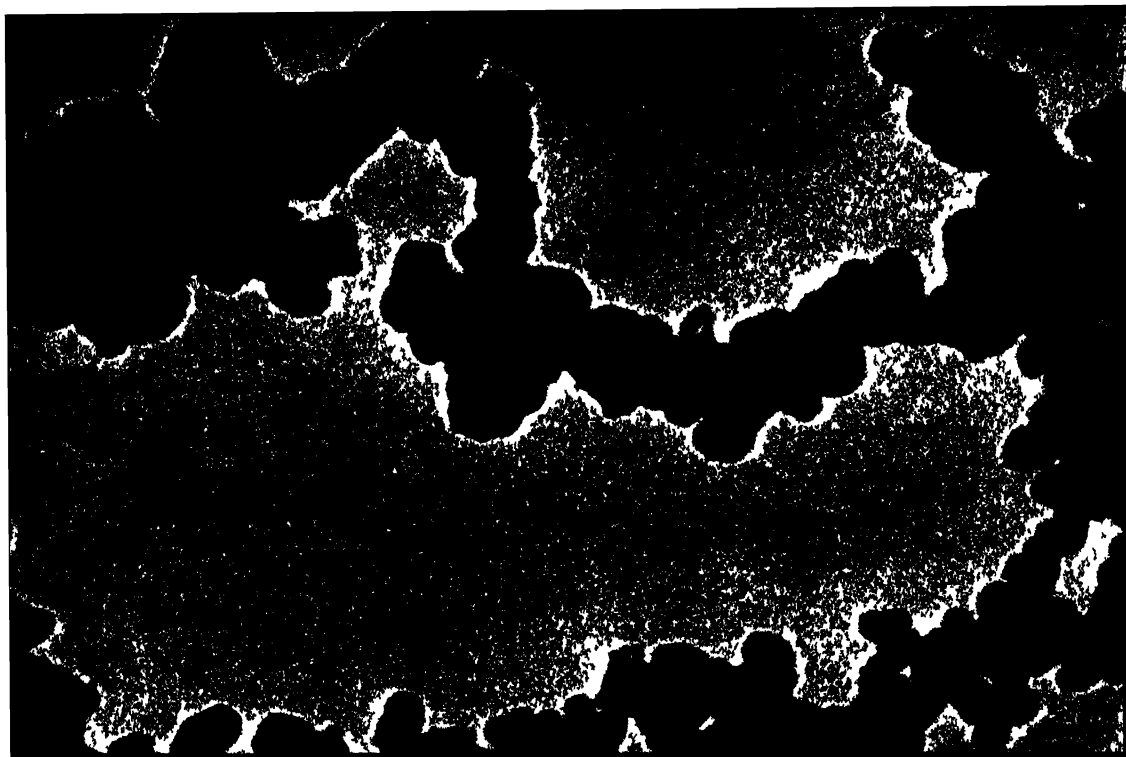


FIG. 6 Enlarged portion of Fig. 4 $\times 100,000$.

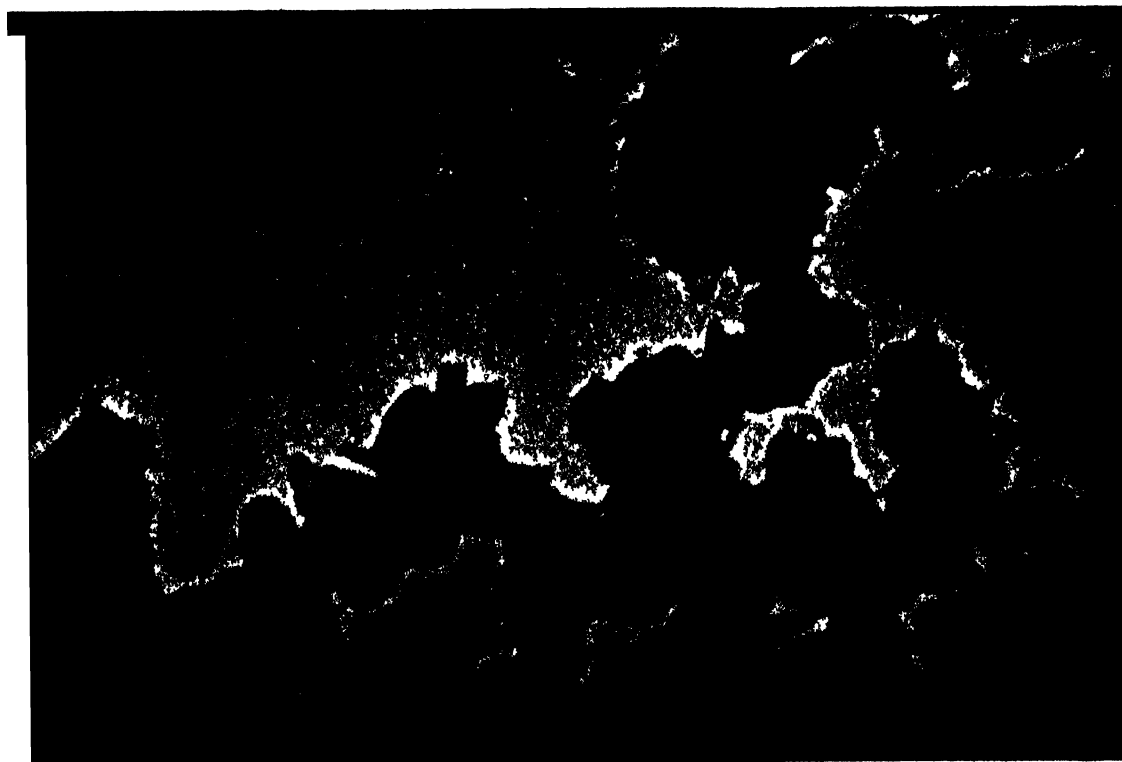


FIG. 7. Enlarged portion of Fig. 5. $\times 100,000$.

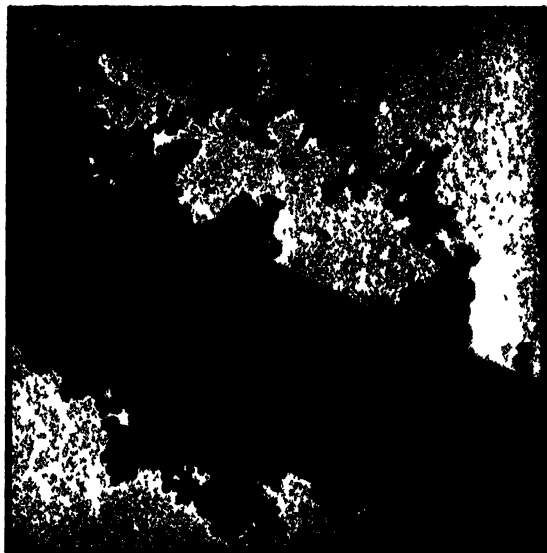


FIG. 8. Glass fiber supporting carbon black. $\times 10,000$ (Shawinigan Chemicals Ltd.).

scattering in these thin crystals is probably of the same order or less than that in the usual thin collodion or Formvar film. Thus, if a specimen of the carbon were mounted upon a film in the usual way, the crystals in the flat position would be practically invisible for lack of contrast. In addition their corners would be rounded off sufficiently to appear as curves owing to lack of resolution. The crystalline appearance would accordingly be lost.

Several observations can be made on the appearance of the carbon specimen in the micrographs. The first is the occurrence of very thin crystal forms at *A* and *B* mentioned above. The second is the presence of much thicker particles than those at *A* or *B* which possess hexagonal shape as at *C*, and the fact that many of these thicker particles have straight-line rather than curve boundaries. Thirdly, it is to be observed that in general the particles have diverse shapes which appear to be oval or spherical as well as angular. As a fourth observation it is to be noted that many of the edge-on crystal formations present a curved appearance as at *D*. Fifthly the average longest dimension of the particles in this sample is about $500 \pm 25\text{\AA}$, and in some cases their thickness may be as low as $30\text{\AA}$.

In order to offer an explanation for the first

observation it seems that a slight difference between the particle structure as observed here and that suggested by Biscoe and Warren⁴ must be advanced. There may be piles of the carbon platelets which are oriented in a symmetric rather than a random fashion about the common perpendicular to the surfaces. This will not only be true of the thinnest particles as at *A* and *B* but will be true of thicker particles as at *C*.

To account for the second observation it is necessary to conclude that quite large, hexagonal carbon platelets (and consequently pseudo-crystals) can occur frequently. If the platelets are arranged symmetrically in the pseudo-crystal and one such pseudo-crystal makes up a particle, then the particle will appear to be hexagonal. On the other hand several such pseudo-crystals arranged at random would give a particle with random shape. This would explain the occurrence of straight-line edges on some of the particles and the general lack of definite particle shape. Lack of resolution would obscure many of these observations and the particles would appear to be spherical or oval. At times a thin pile of platelets projecting from the main mass can be noted and the crystalline nature recognized.

The fourth observation would suggest that the carbon scales piling up to form the pseudo-crystal do not always build up to the same depth over the whole surface. The resolution however is not such as to render the carbon "steps" visible, and a curved appearance results.

These observations tend to support the contentions of Schoon and Koch⁵ and Von Otto Ruff.⁶ They also follow to a large extent the conclusions of Biscoe and Warren and offer a picture of carbon particle structure which is more compatible with x-ray and electron diffraction evidence than the idea of carbon black as a pure sphere or oval.

This crystalline appearance has been noted so far only in *acetylene* carbon blacks and to the greatest extent in those produced by a method of thermal decomposition. The occurrence of this

⁴ J. Biscoe and B. E. Warren, J. App. Phys. 13, 364 (1942).

⁵ Th. Schoon and H. W. Koch, Kautschuk (January, 1941).

⁶ Von Otto Ruff, Zeits. f. Elektrochem. angew. physik. Chem. 44, 333 (1938).

crystalline habit may account not only for the higher conductivity of the acetylene blacks but may be a factor also in the formation of their strong structure. However, it is to be noted here that all acetylene blacks, produced by different processes, do not have the same appearance in electron micrographs nor are they at all identical in other properties. Acetylene black as such has been compared with blacks formed from other raw products. Such comparisons have little

meaning unless the method of preparation of the acetylene black is clearly specified.

It is hoped that this mounting technique or adaptations of it will prove useful in further investigations of this and other carbon blacks. The author is indebted to the Plant Research Department of the Shawinigan Chemicals Limited and to the Physics Department of the University of Toronto for permission to reproduce the micrographs.

Investigation of Secondary Phases in Alloys by Electron Diffraction and the Electron Microscope

R. D. HEIDENREICH,* L. STURKEY, AND H. L. WOODS
The Dow Chemical Company, Midland, Michigan

(Received October 15, 1945)

The application of electron diffraction reflection methods and the electron microscope to investigations of metallurgical structures is reported. It is shown that these instruments can yield valuable information concerning very small amounts of secondary phases and impurities in metals. Thus, for example, when martensite decomposes at 200°C, the slow etching component is highly dispersed Fe_3N . The decomposition at 400°C yields the usual Fe_3C . Further, the Fe_3N formed at 200°C transforms to Fe_3C when annealed at 350°C. Other examples are presented. The successful application of these instruments is dependent upon the production of clean, etched surfaces. Surface preparation is discussed and a satisfactory technique for rinsing etched surfaces is described. In some alloy systems, the etching reaction may result in the redeposition of a component onto the etched surface. In the case of a β -brass containing 0.0005–0.001 percent silver, the silver is deposited in the form of spherical particles about 1000Å in diameter. Aluminum alloys appear to offer a serious problem in discovering satisfactory etchants due to the position of aluminum in the electromotive series relative to copper, silver, etc.

INTRODUCTION

THE physical and chemical properties of metallic alloys are well known to be in a large measure dependent upon the distribution and nature of a secondary constituent in the crystals of the base metal. The detection and identification of these secondary phases and structures is often difficult and tedious using x-ray and light microscopic methods and in many instances is nearly impossible when the concentration of such phases is very small.

In x-ray studies of alloys where information concerning precipitated phases is desired, the energy diffracted by the minor phase may constitute only a small fraction of that diffracted by

the matrix crystals. If this precipitate is very highly dispersed or amounts to less than about 5 percent of the total diffracting material it may not contribute to the final pattern.

It has been found that by using the electron microscope and a refined electron diffraction reflection method in close conjunction, information concerning very small amounts of secondary phases and impurities can be readily obtained. If an alloy is etched so that the minor constituents are left protruding above the matrix crystals (provided no extraneous material is deposited on the surface), a very small amount of included material can yield a strong electron diffraction pattern without an interfering pattern of the matrix.

In the past, electron diffraction has been

* Now at Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

applied chiefly to studies of polishing and grinding and to the structure of corrosion films, etc. There is reason to believe that with suitable surface preparation methods, electron diffraction techniques may be of great value in investigations where a minor phase plays an important role.

The success of this method of investigation of metallurgical structures is dependent upon the production of clean, etched surfaces. The usual metallographic techniques result in surfaces that bear oxide or hydroxide films or reaction products remaining on the surface after etching.

The first part of this paper will be concerned with obtaining satisfactory surfaces while the remainder will deal with some of the results obtained with several alloys.

SURFACE PREPARATION

In preparing metal surfaces for combined investigation by electron diffraction and electron microscopy several considerations are of major importance:

(1) The removal of all distorted and fragmented metal in the surface layers produced during grinding and polishing.

(2) The reduction of the quantity of corrosion products remaining on the etched surface below the minimum necessary to produce appreciable electron scattering.

(3) Selectivity of the etchant for the various metallic constituents so that minor phases are left standing in relief on the surface.

The distortion of the surface regions by the usual metallographic preparations has been investigated by means of x-ray¹ and electron diffraction studies.² Electron microscope evidence³ in the case of rocksalt is consistent with the assumption of fragmentation and of the formation of a polish or Beilby layer during mechanical polishing. Optical investigations⁴ with polished copper likewise bear out these conclusions. In view of these phenomena it was concluded that mechanical grinding and polishing introduced

grave complications in this type of investigation if sufficient metal were not removed chemically during etching. This objection is readily overcome by the use of electrolytic polishing methods.

As a result, the procedure consisted in abrading the metal specimens through 4/0 metallographic paper and then electrolytically polishing with no further mechanical disturbance. Electrolytes and conditions for most metals have been published.⁵ Electrolytically polished surfaces often tend to be wavy and not very satisfactory for light microscopy. For electron microscopy, however, such surfaces are quite acceptable due to the great depth of focus of the electron microscope and the fact that the area of the field viewed is small compared to the ripples that may occur in the polished surface.

After polishing, the metal surface is etched by a selective chemical reagent. The type and rate of attack of a metal crystal are determined by several factors such as the reagent used, temperature, concentration, phases present, and the behavior of the reaction products. The deposition or adsorption of reaction products or of a specific material present in the reagent may greatly alter the type of attack.^{6,7}

The most critical step in obtaining clean surfaces is that of immediate rinsing after removal of the specimen from the etching solution. A sequence of rinses has been worked out whereby the etching solution is finally displaced by an inert liquid without deposition of foreign materials. For several alloys, the following sequence yields clean surfaces after etching:

```

Etchant
↓
Rinse I  50:50 methanol acetone with ½ to
          1 percent formic acid
↓
Rinse II  Methanol-acetone
↓
Rinse III Clean, dry benzene. Dry in blast
           of clean, warm air.
  
```

The success of the rinsing procedure for various metals depends upon the use of rinses in which

¹ H. C. Vacher, J. Research Nat. Bur. Stand. 29, 177 (1942).

² G. P. Thomson and W. Cockrane, *Theory and Practice of Electron Diffraction* (The Macmillan Company, New York, 1939), Chap. XIII, see also, Reference 9.

³ R. D. Heidenreich, J. Opt. Soc. Am. 35, 139 (1945).

⁴ H. Lowry, H. Wilkonson, and D. L. Smare, Phil. Mag. 22, 769 (1936).

⁵ For a review see O. Smeckal, Metal Progress, 729 (April, 1945).

⁶ C. H. Desch, *Chemistry of Solids* (Cornell University Press, Ithaca, 1934), Chap. V.

⁷ H. Mahl and I. N. Stranski, Zeits. f. physik. Chemie 51, 319 (1942).

any metallic compounds formed are soluble and are hence not precipitated onto the surface. In all cases, clean, freshly distilled solvents should be employed.

The results are not independent of the etching reagent used. Reagents which stain the surfaces are to be avoided as well as those which tend to passivate or form insoluble films. For example, in the case of steel specimens, nitric acid is not satisfactory probably due to a thin film (amorphous) formed during etching which partially obscures the underlying metal and yields poor diffraction patterns. Ferric chloride in hydrochloric acid, on the other hand, gives excellent results.

It will be found that some alloy systems will require considerable investigation in order to discover the proper etching reagent since the common reagents may result in one or more of the constituents being dissolved and then re-deposited on the surface in a different form.⁸ This is particularly true of aluminum-copper alloys and impurities in several copper alloys as will be shown.

EXAMINATION OF SURFACES

The electron microscope examination of the etched surfaces was carried out using the polystyrene-silica replica method after electron diffraction reflection patterns had been obtained from the surfaces.⁹ The necessity of employing a replica is a disadvantage but, nevertheless, the replicas and electron diffraction patterns make it possible to deduce considerable information concerning the surface structures. All micrographs were taken with an RCA, Type B electron microscope. The electron diffraction apparatus employed in this work had a 75-cm plate-specimen distance and a well-collimated parallel incident beam of 40 kv electrons.

The relation between the electron diffraction patterns and the electron micrographs is of immediate interest. The diffraction patterns are determined by the crystal structure and orientation of those surface protuberances which exhibit a thickness normal to the incident beam sufficiently small to allow transmission. Thus, corners,

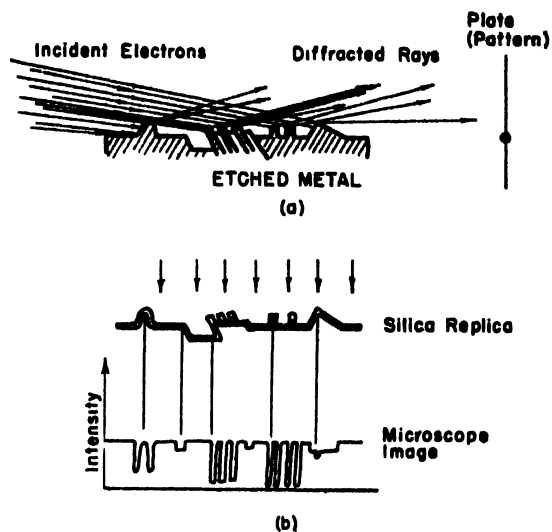


FIG. 1. Effect of surface contours on an electron diffraction reflection pattern (a) and electron microscope image of a surface replica (b). Only the diffracted beams are shown leaving the surface in (a).

edges, and thin peaks that can be seen in the electron micrographs are to be considered as possible diffracting regions of the surface. Figure 1 illustrates the approximate relation between the image and the diffraction pattern based on the topography of an etched surface.

The interpretation⁹ of electron diffraction patterns which, on the basis of x-ray experience, would indicate preferred orientation is to be made with particular caution. In many cases, the crystallographic etch attack results in certain crystal faces having a topography such that they do not contribute to the diffraction pattern. This point has been discussed by Johnson and Gram¹⁰ for etched tungsten and molybdenum and is thought by the authors to be a general consideration in the case of pure metals. Two-phase systems are not so likely to give difficulty.

EXAMPLES

Several alloy systems to which the surface preparation methods outlined have been applied will serve as examples of the kind of information that can be obtained. The illustrations each con-

⁸ Cf. J. H. Hollomon and J. Wulff, *Trans. A.I.M.E.* **143**, 93 (1941).

⁹ Cf. general discussion on "Structure of metallic coatings, films and surfaces," *Trans. Faraday Soc.* **31**, 1043 (1935).

¹⁰ R. P. Johnson and W. R. Gram, *Phys. Rev.* **62**, 77 (1942).

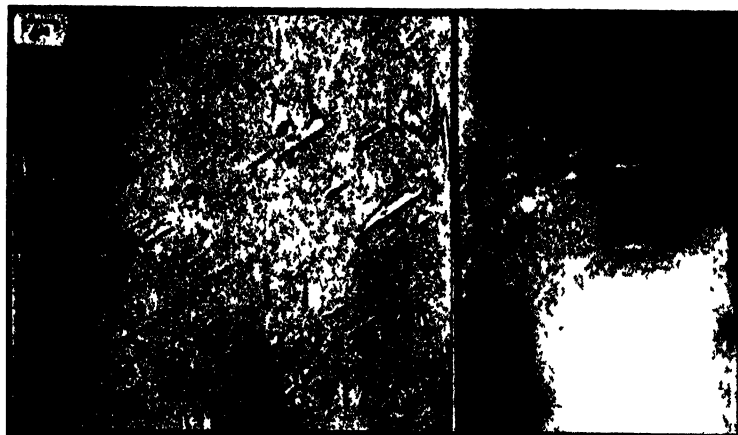


FIG. 2. Electron micrograph and diffraction pattern from magnesium single crystal containing 1.5 percent manganese. The manganese particles are plate-shaped and are preferentially oriented in the magnesium crystal.

sist of an electron micrograph (silica replica) and sometimes an electron diffraction reflection pattern from the same surface. Reflection patterns are often difficult to reproduce so that the rings are easily visible. Measurements of the radii of the diffraction rings are made on the original negatives and in some instances tables of d_{hkl} values are included for comparison with x-ray standards. Identification is generally accomplished by such a comparison.

MAGNESIUM ALLOYS

Although this work was initiated by research with magnesium alloys, the etching methods and results will not be included here. The technique for magnesium is sufficiently unique and the results extensive enough to warrant a full paper devoted to this subject alone which will appear elsewhere. One example will suffice.

Figure 2 shows an electron micrograph and electron diffraction reflection pattern from an etched single crystal of magnesium containing 1.5 percent manganese. The manganese is in the form of plate-shaped particles precipitated along certain crystallographic planes* of the crystal. The individual particles cannot be resolved with a good light microscope and hence their shape and arrangement cannot be determined. The electron diffraction pattern in Fig. 2 is that for α -Mn with preferred orientation. The preferred orientation indicated by this pattern is not a result of preferential etching but is

* The etched surface is the basal plane and the particles are oriented along the three principle directions.

indicative of the relation between the particles and the matrix crystal

STEEL

The results obtained with a plain carbon steel have been briefly reported.¹¹ They are of considerable interest and value both from a theoretical and practical viewpoint. The additional information obtainable by combined electron diffraction and electron microscope studies on properly prepared surfaces is clearly illustrated.

The steel employed in this series of heat treatments analyzed 0.9 percent C, 0.15 percent Si, and 0.0056 percent N for the major constituents. The surfaces were prepared by abrading through 4/0 metallographic paper, electrolytically polishing in sulfuric-phosphoric acids,¹² and then etching in a methanol solution containing 2 percent HCl and 1 percent FeCl_3 . The etching times were 30 sec. for pearlitic structures and 60–90 sec. for martensitic specimens. The standard sequence of rinses enumerated in a previous section were employed. As stated, nitric acid etchants are not recommended and the use of picric acid is questionable.

The results for this steel when heat treated to produce fine pearlite are seen in Fig. 3.

The micrograph shows the usual lamellar structure and the electron diffraction pattern is that for polycrystalline, orthorhombic Fe_3C . The pattern for Fe_3C is to be expected since the

¹¹ R. D. Heidenreich, L. Sturkey, and H. L. Woods, *Nature* (in press).

¹² R. I. Imboden and R. S. Sibley, *Metal Finishing*, 592 (November, 1942).

etchant attacks the α -Fe more rapidly, leaving the cementite lamellae standing in relief. For the sake of completeness, Table I lists the d_{hkl} values for Fe_3C compared with values obtained from measurements on the electron diffraction pattern.

The structure of pearlite is much better understood than is that of martensite and for that reason martensite is of greater interest here. Martensite is the body-centered tetragonal product obtained when austenite transforms at low temperatures and is produced by quenching the steel from 850°C into ice water. Martensite is quite unstable and decomposes at temperatures as low as 100°C . The decomposition products are of special interest since they apparently are dependent upon the temperature at which the reaction proceeds.¹³ It was generally thought that decomposition below 300°C yielded a highly dispersed carbide phase, the particles of which were sufficiently strained by the stresses set up during precipitation that no satisfactory x-ray pattern could be obtained.¹⁴ Above 300°C , the product was established to be Fe_3C by x-ray methods.

The surface of martensite, etched and rinsed as described, is seen in Fig. 4. The appearance

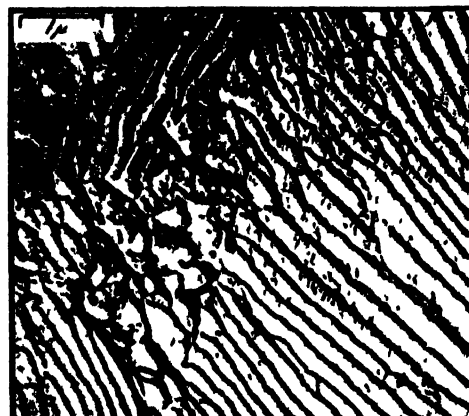


FIG. 3. Electron micrograph for fine pearlite in a 0.9 percent carbon steel etched in $\text{HCl}-\text{FeCl}_3$. Rockwell C. 29.

¹³ M. Arbusow and G. Kurdjumow, J. Phys. U.S.S.R. 5, 2-3, 101 (1941).

¹⁴ C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 479 (further references will be found here).

TABLE I. Comparison of electron diffraction results from sample of Fig. 3 with x-ray standard of Fe_3C .

{hkl}	d-values		Relative intensities	
	x-ray*	ED	x-ray	ED ^b
112	2.39	2.41	medium	
200	2.265		weak	
121	2.115		medium	
210	2.07		medium	
022	2.04	2.08	strong	very strong
211	1.98	1.99	medium	
113	1.875	1.86	strong	strong
212	1.766		weak	
004	1.69		weak	
221	1.645		very weak	
130	1.593	1.595	weak	
131	1.55		very weak	
114	1.512		weak	
311	1.415		weak	
231	1.335	1.34	medium	medium strong
015	1.31		very weak	
232	1.262		very weak	
140	1.23		medium	
313	1.22	1.22	medium	
233	1.165	1.16	strong	strong
142	1.156		strong	
400	1.133		strong	
006	1.127	1.125	strong	
331, 240	1.112	1.102	strong	
242	1.058		medium	
413	0.995		medium	
(150, 216)	0.987	0.980	strong	

* X-ray values taken from A. Westgren and G. Phragmen, J. Iron Steel Inst. 105, 241 (1922).

^b ED = Electron diffraction reflection

seems to be consistent with the present theory of martensite formation.¹⁵ The dark regions in the micrograph of Fig. 4 should appear down and are probably due to untransformed austenite. The diffraction pattern is not only weak, but the rings are broader and more diffuse than generally found by x-ray diffraction.¹⁶ This can be taken to indicate a crystal with heterogeneous internal stresses exhibiting no sharply defined lattice spacings but, rather, d -values fluctuating about an average. (See Table II.)

When the specimen of Fig. 4 is annealed $\frac{1}{2}$ hr. at 200°C , the martensite plates decompose into the colloidal mixture seen in Fig. 5. The slow-etching component is not Fe_3C as often assumed but is clearly identified with hexagonal Fe_3N from the reflection pattern. Table III lists the d_{hkl} values. Further heating for 4 hr. at 200°C does not change the appearance, diffraction pattern, or hardness.

¹⁵ Reference 14, p. 483.

¹⁶ E. Ohman, J. Iron Steel Inst. 123, 445 (1931).



FIG. 4. Results on the same steel as Fig. 3 after heating at 850°C and quenching into ice water where the transformation austenite→martensite occurs. The diffraction pattern is weak and diffuse but is characteristic of martensite. Etched in HCl-FeCl₃. Rockwell C. 64.

TABLE II. Electron diffraction results from etched martensite compared with x-ray standard of α -Fe.

(hkl)	d_{hkl}		Relative intensity	
	x ray	ED	x-ray	ED
110	2.01A	2.0-2.1A	40	Broad, diffuse rings
200	1.425	1.4-1.5	8	
211	1.164	1.1 1.2	15	
220	1.508		4	
310	0.902	0.9	2	
222	0.823		1	
321	0.763		3	
411	0.677		1	

• ED = Electron diffraction reflection.



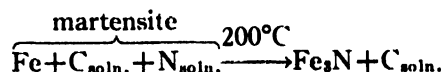
FIG. 5. Decomposition of martensite by annealing $\frac{1}{2}$ hour at 200°C. The slow etching phase is Fe₃N (or isomorphous with Fe₃N). Etched in HCl-FeCl₃. Rockwell C. 58.

TABLE III. Electron diffraction results from sample of Fig. 5 compared with x-ray standard of Fe₃N.

(hkl)	d_{hkl}		Relative intensity	
	x-ray	ED*	x-ray	ED
020, 110	2.38A		8	
002	2.19	2.18A	10	Strong
111, 021	2.09	2.09	40	Very strong
112, 022	1.61	1.61	10	Strong
130, 200	1.375	1.38	10	Strong
	1.240	1.23	10	Strong
	1.160	1.157	8	Strong
	1.140		4	Weak
	1.095	1.10	1	Weak

* ED = Electron diffraction reflection.

The decomposition reaction at 200°C has been suggested¹¹ to be



The nitrogen analysis gives some idea as to the small amount of the nitride phase that would be present. The greater part of the carbon must reside in α -Fe, probably as solid solution.

If the specimen of Fig. 4 is annealed $\frac{1}{2}$ hr. at 400°C, the result is that of Fig. 6. A colloidal mixture is again produced but the slow etching component is now Fe₃C as ascertained by electron diffraction. When the specimen of Fig. 5 is further heated $\frac{1}{2}$ hr. at 350°C, the patterns indicate that a mixture of Fe₃N and Fe₃C is produced. The appearance is illustrated in Fig. 7.

The decomposition reaction at 400°C is suggested to be

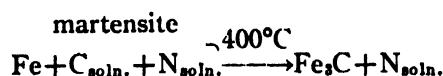
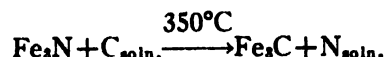


Figure 7 indicates that the following reaction also may occur



It appears from these preliminary results that the observations of Arbusow and Kurdjumow¹² are accounted for by the formation of Fe₃C above 300°C and the formation of Fe₃N below 300°C. The behavior of steel specimens from which the nitrogen has been thoroughly removed would be of great interest. The possibility of a

carbide phase isomorphous with Fe_3N should not be ignored.

The hardness of this series of specimens was determined and the values can be crudely correlated with the dimensions of the "particles" or "blocks" formed in the decomposition. The hardness is plotted against "block size" in Fig. 8 although caution should be used in drawing any general conclusions from the data. The result does seem to be consistent with the observations of Barrett¹⁷ and may bear an interesting relation to W. L. Bragg's¹⁸ theory of the strength of metals.

ALUMINUM ALLOYS

Preliminary experiments on the etching of an aluminum-copper alloy indicate that consider-

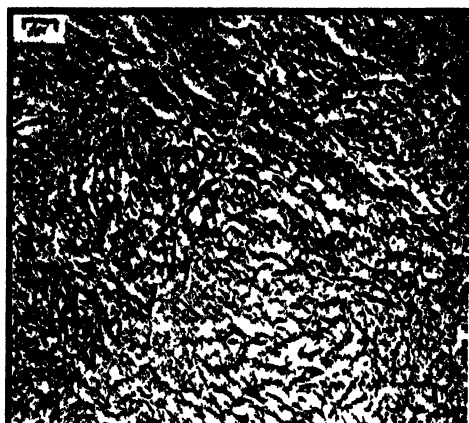


FIG. 6. Decomposition of martensite by annealing $\frac{1}{2}$ hr. at 400°C. The slow etching phase is now the usual Fe_3C . Etched in $\text{HCl}-\text{FeCl}_3$. Rockwell C. 47.

able work may be required in order to find a suitable etchant. Pure aluminum can be etched in hydrochloric acid and rinsed as with magnesium or steel and will yield good diffraction patterns for Al. An aluminum-copper alloy, however, when etched with the common reagents (HCl , HF , HNO_3 , and mixtures) does not produce the desired result. In all cases the diffraction pattern obtained was that for polycrystalline

¹⁷ C. S. Barrett, *Metals Tech.* 10 (Sept. 1943).

¹⁸ W. L. Bragg, *Nature* 149, 511 (1942).

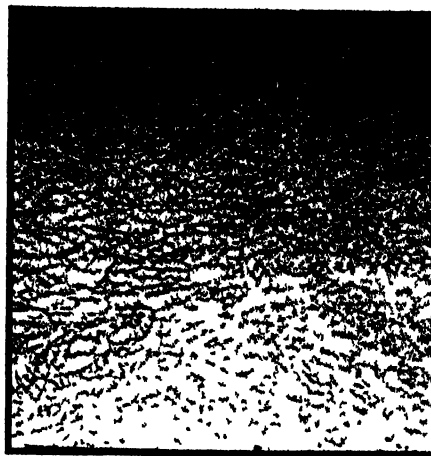


FIG. 7. Transformation of Fe_3N (formed by heating martensite $\frac{1}{2}$ hour at 200°C) to Fe_3C by further heating $\frac{1}{2}$ hour at 350°C. Etched in $\text{HCl}-\text{FeCl}_3$. Rockwell C. 52.

copper with sometimes a trace of CuO . It seems likely that as both the aluminum and copper dissolved in the reagent the copper was re-deposited as a result of the relative positions of the two metals in the electromotive series. This behavior produces a surface which may have little or no resemblance to the structure of the bulk metal and so makes the combined diffraction and microscopic technique of little value from a strictly metallurgical viewpoint. From the standpoint of corrosion studies and film formation, numerous experiments suggest themselves wherein information can be obtained using these methods.

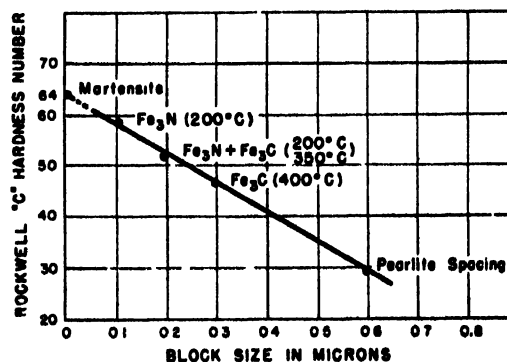


FIG. 8. Rough correlation between "particle" or "block" dimensions in heat treated steel with Rockwell C hardness values.

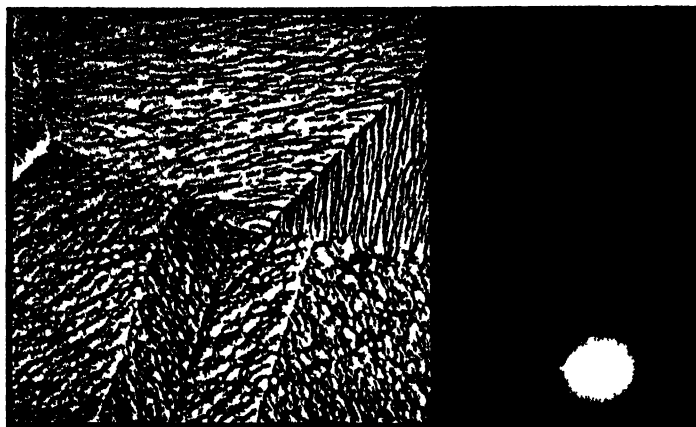


FIG. 9. Electron micrograph and diffraction pattern from commercial copper sheet etched in ammonium persulfate. The pattern is that for pure copper and exhibits arcing. The arcing is attributed to the etch attack rather than the orientation of the sheet.

COPPER ALLOYS

Inasmuch as the electron microscope and electron diffraction constitute sensitive techniques of investigating minute amounts of new phases, it would seem natural to apply them to studies of age-hardening alloys. Beryllium copper is an excellent example of an age-hardening alloy, and the work on copper was pointed toward an investigation of this system. Only the first experiments have been carried out, however.

There are several common etchants for copper and its alloys. Ammonium persulfate is satisfactory for pure copper and produces surfaces that yield strong diffraction patterns. The preparation for the commercial sheet shown in Fig. 9 consisted in electrolytically polishing in a phosphoric acid electrolyte, etching in a 10 percent solution of ammonium persulfate in distilled water, rinsing in distilled water and then through rinses I, II, and III. This surface is very rough on a micro-scale and for this reason possesses many sharp peaks that serve as diffracting centers. The reason for the etch structure of Fig. 9 has been briefly discussed by Barrett.¹⁹ It would appear that the copper grains consist of a mosaic of rods about one micron long and a few tenths of a micron in diameter. It is evident that the orientation of the rods is related to specific directions in the copper crystal. The dimensions of the "etch blocks" are not independent of the

etchant employed and therefore should be viewed cautiously until the matter has been well investigated.

The diffraction pattern of Fig. 9 shows the rings breaking into spots due to the grain size of the sample. In addition, there is arcing of the rings which was first thought to indicate preferred orientation as might be expected in rolled sheet. When this same sheet was deformed and cold

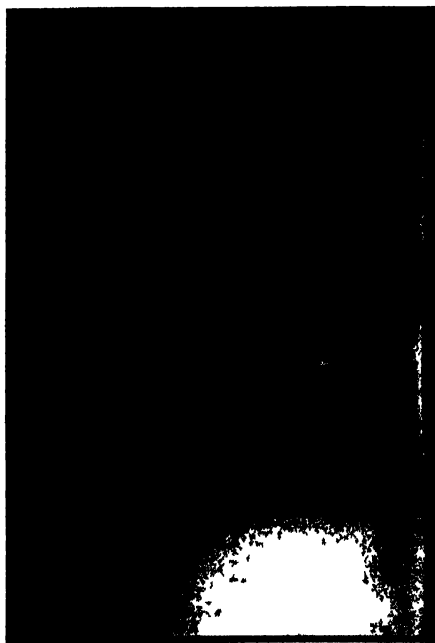


FIG. 10. Electron diffraction pattern of the same copper sheet as Fig. 9 after cold working and annealing. The grain size has increased but the arcing is still about the same as in Fig. 9.

¹⁹ C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 223.



FIG. 11. Results for β -brass containing 0.0005-0.001 percent silver. The pattern obtained is that for polycrystalline silver, the particles of which can be seen in the micrograph as a result of being transferred to the silica replica. Etched in acetic-nitric reagent from which zinc displaces silver.

worked and then annealed, the pattern of Fig. 10 was obtained. The grain size has increased but the arcing is about the same as in Fig. 9. It appears that the etch attack is the chief factor in determining the arcing of the rings as previously discussed.

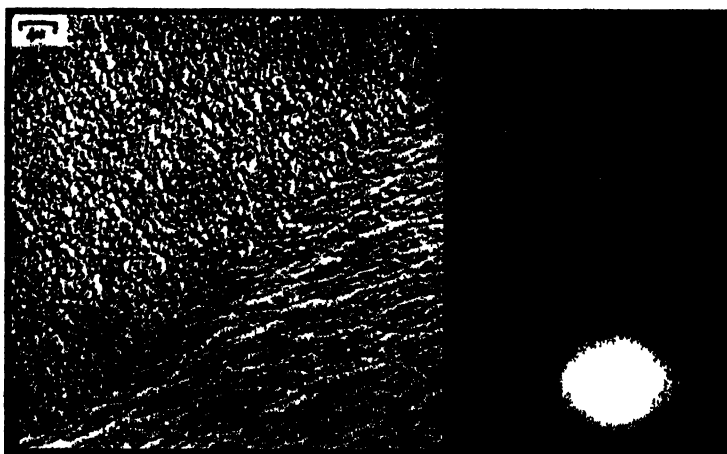
With certain reagents, metallic impurities in copper and brass may be deposited on the surface as shown by Holomon and Wulff.⁸ An example of this is the β -brass of Fig. 11. This sample was electrolytically polished, etched in glacial acetic-nitric acids (5:2), rinsed in dilute ammonium hydroxide, distilled water, and then

the usual rinses. This etchant was employed because it did not tend to stain the surfaces. The diffraction pattern from the etched surface was that for metallic silver, however. When the silica replica was examined it was found to be contaminated with black particles which had been transferred from the original surface to the replica. A spectroscopic analysis of the original sample indicated 0.0005 to 0.001 percent Ag from which it was concluded that, during etching, colloidal silver particles about 1000Å in diameter had been deposited on the surface. A subsequent β -brass made from electrolytic copper and high purity zinc did not present this difficulty but gave a good pattern for body-centered brass. The transfer of foreign particles from a metal surface to the silica replica has been observed many times.

The final example, beryllium copper, has not been subjected to a systematic investigation due to two difficulties. The first of these is the deposition of an impurity metal (less than 0.01 percent) during etching in the acetic-nitric reagent. This problem is not encountered in the use of ammonium persulfate but the type of etch structure makes it difficult to follow the aging process. The second difficulty has been the failure to produce a complete solid solution of beryllium in copper. A Cu-2 percent Be alloy was chill cast and then solution heat treated 100 hr. at 820°C and quenched.* The diffrac-

* Specimens $\frac{1}{8}$ " in diameter and $\frac{1}{16}$ " thick.

FIG. 12 Electron micrograph and diffraction pattern for beryllium copper (2 percent Be) quenched after 100 hours at 820°C. The spots are due to the copper solid solution while the rings are ascribed to both Cu and the intermetallic compound, CuBe, which has not entirely dissolved upon solution heat treatment. Etched in ammonium persulfate.



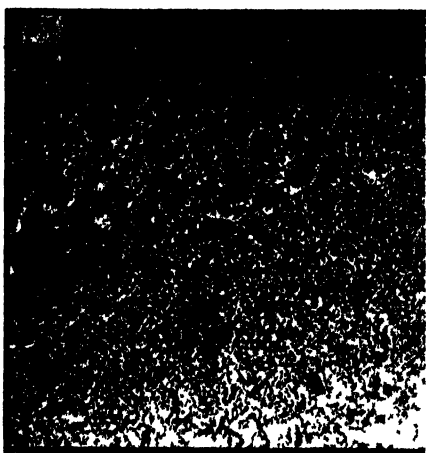


FIG. 13. Results for the same specimen as Fig. 12 after annealing 24 hours at 300°C. The compound is now densely precipitated and its diffraction pattern is that for CuBe with a trace of copper due to the matrix.

tion patterns from the surface etched with ammonium persulfate show not only the expected

pattern (spots) for copper but also several rings due to the compound CuBe. Fig. 12 shows the micrograph and diffraction pattern.

If the specimen that yielded the results of Fig. 12 is aged 24 hr. at 300°C in order to heavily precipitate the compound, the result is that of Fig. 13. The particles of CuBe are evident in the micrograph. The pattern is that for polycrystalline CuBe with a trace of copper, as would be expected. The mechanism of age hardening²⁰ in this alloy cannot be studied, however, until a solid solution completely free of CuBe is attained. When such a solid solution is produced, a sequence of samples annealed for various lengths of time should yield some valuable information on which to base a theoretical study of age hardening.

²⁰ See, A. Guinier and P. Jacquet, *Nature* **155**, 695 (1945).

Journal of Applied Physics

Volume 17, Number 3

March, 1946

Film-Lubrication between Spherical Surfaces: with an Application to the Theory of the Four-Ball Lubricant Testing Instrument

J. HOWLETT

Physics Department, University of Manchester, Manchester, England, and Research Department, London, Midland and Scottish Railway Company, Ltd., Derby, England

(Received July 19, 1945)

The paper is concerned mainly with the flow of viscous liquid between the outer surfaces of two equal spheres placed in an infinite body of the liquid so that the "minimum separation" of the surfaces (the distance apart at the points of closest approach) is small compared with the radius; one sphere is held at rest, the other rotates at constant speed about an axis which does not pass through the center of the fixed sphere. The assumption is made that the surfaces are separated by a continuous film of liquid to which the classical theory of hydrodynamical lubrication can be applied—that is, "boundary lubrication" and actual contact between the surfaces are excluded; the partial differential equation (the Reynolds equation) for the pressure distribution in this film is set up, and it is shown that with a suitable choice of boundary conditions a solution can be found in very simple form. From this solution are deduced the total force exerted by the liquid on the moving sphere and the torque, about the axis of rotation, of the forces acting on the fixed sphere. The results are applied to the particular case of the four-ball lubricant testing instrument, in which a sphere is made to rotate about a vertical axis, under axial load, in the central space formed by a set of three equal spheres held stationary so as to touch one another with their centers in a horizontal plane; the whole set of spheres is immersed in the lubricant to be tested. The relations finally obtained express the load

on the moving sphere and the torque on the set of fixed spheres in terms of the radius of the spheres, the velocity of rotation, and the viscosity of the liquid; they involve as a parameter the minimum separation of the moving sphere from the fixed spheres, and elimination of this gives the relation between the measurable quantities, load, and torque, which could provide a basis for the use of the instrument as a viscometer. It is found that the torque is a very slowly varying function of the load, being approximately proportional to the logarithm of the load. Further, if it is assumed that the hydrodynamical theory of lubrication is valid only for films whose thickness is greater than some agreed minimum, then the expression for the load on the moving sphere, into which the minimum separation enters, provides a criterion for the range of conditions (of which the load is much the most important) within which the theory does in fact apply to the instrument. Numerical examples are considered, and it is shown that, with allowance for wide departures from the standard size and speed of operation of the instrument, the maximum load which can be supported by the moving sphere in the hydrodynamical regime cannot be expected to exceed a few hundred grams weight; for loads of greater order the standard continuous-film theory cannot hold, and the forces on the spheres are no longer determined by the viscosity of the liquid.

INTRODUCTION

IN the last few years a great deal of attention has been given to the properties of lubricants under conditions of "extreme pressure," that is,

under pressures so great (of the order of 10^4 kg wt./cm²) that it is not to be expected that the lubricated surfaces are separated by the continuous film which characterizes hydrodynamical

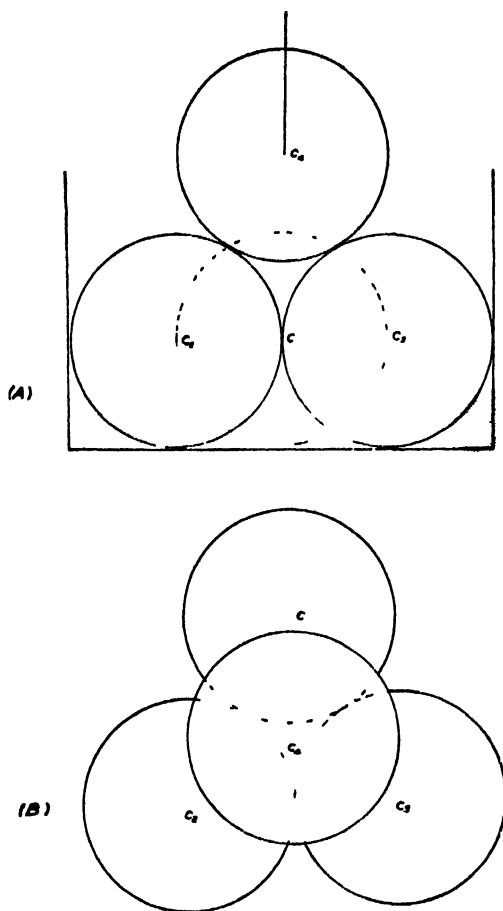


FIG. 1. (A) Elevation and (B) plan of the 4-ball instrument. C_1 , C_2 , C_3 are the centers of the stationary spheres and C_4 that of the rotating sphere.

or "thick film" lubrication. The literature of the subject has many references to the Boerlage four-ball testing instrument.¹

The instrument seems to have become a standard piece of apparatus in this field, but at the same time it seems that the interpretation of the measurements made with it is not always certain. There is an obvious question which must present itself in any general discussion of the theory of the apparatus, and which should be amenable to theoretical investigation; this is, can the instrument behave in any circumstances according to the laws of hydrodynamical lubrication, and if it can, what is the range of operating conditions over which this is possible? There seems to be no

published investigation of this point; the present paper attempts a treatment based only on classical hydrodynamical theory.

We may recall here that the four-ball instrument consists essentially (see Fig. 1) of a set of three equal spheres held rigidly in a fixed cup so as to touch one another with their centers in a horizontal plane, with a fourth sphere rotated under load about a vertical axis in the central space of the fixed spheres; the whole is immersed in the lubricant to be tested. Details of some methods of use with great loads (which are not really relevant to our present purpose) are given in other papers.¹ We base this investigation on two fundamental assumptions; first that the spheres are not deformed by the load, and second, that the flow of liquid around and between the spheres obeys the classical equations of viscous flow, in particular that a continuous film in the hydrodynamical sense (as distinct from the "boundary lubrication" sense) is formed between the moving sphere and each fixed sphere.

We attack the problem indirectly by first developing the relation between the thickness of this film at its thinnest part (the "minimum film thickness," say) and the load applied to the axis. Suppose we find a relation $W = W(h_0)$ between the load W and this minimum thickness h_0 —involving of course all the parameters of the problem—the radius of the spheres, the angular velocity of the rotating sphere, and the viscosity of the liquid. Suppose also that we have reason to believe that a film of thickness less than some value d does not behave as a lubricating film in the hydrodynamical sense; then we may deduce that the assumption of hydrodynamical lubrication ceases to hold for loads exceeding the value $W(d)$ corresponding to the thickness $h_0 = d$. Conversely we may deduce that for loads less than this value there will be in fact lubrication of the hydrodynamical type. We have thus a criterion for the range of application of the classical theory.

A complete and rigorous investigation of the problem would involve the solving of the full equations for the flow around the spheres, and with a geometry so complicated no success could be expected from such an undertaking; indeed it was shown by Lamb² that even the apparently

¹ G. D. Boerlage, *Engineering* 136, 46 (1933). O. Beek, *J. App. Phys.* 12, 512 (1941). D. Clayton, *Inst. Mech. Eng., Discussion on Lubrication* 2, 282 (1937). R. Schnurmann, *Engineering* 149, 567 (1940).

² H. Lamb, *Hydrodynamics* (Cambridge University Press, London, 1932), p. 589.

simple problem of a sphere rotating in an infinite body of liquid is intractable if the velocity exceeds a very small value. To bring the problem within the scope of reasonably simple analysis we make the following further assumptions:

A. The flow in the immediate vicinity of the moving sphere and any one of the fixed spheres is unaffected by the presence of the other spheres.

B. The film formed between the moving sphere and any fixed sphere is sufficiently thin for the standard assumptions of classical film-theory to hold; that is,

- (a) the inertia terms in the equations of motion can be neglected,
- (b) the pressure can be regarded as constant across the film, and
- (c) rates of change of velocity measured across the film are much greater than those measured in any other direction.

There is also the over-riding assumption, implied in the use of the hydrodynamical equations, that the only relevant physical property of the lubricant is its viscosity; and we shall take this as constant throughout the film.

With these assumptions the problem is reduced to the solving of the Reynolds equation for the flow between one pair of spherical surfaces, that is, the partial differential equation for the pressure distribution in the film. The general outline of the investigation is then as follows: First (§§1, 2) the Reynolds equation is set up and the boundary conditions are obtained. It is then shown that with a suitable choice of boundary conditions there is a very simple solution which is computed as a numerical function. From this we deduce (§4) the total load supported by the film; and a further analysis (§5) leads to the expression for the viscous-frictional torque exerted on the stationary sphere. Finally we apply these results to the particular case of the four-ball instrument and derive the criterion for the range of working within which we may suppose hydrodynamical lubrication to occur.

The final conclusion is that the usual conditions of heavy loads are far outside this range, as is to be expected, so that in these circumstances the relevant property of the lubricant is not its viscosity. The results have other points

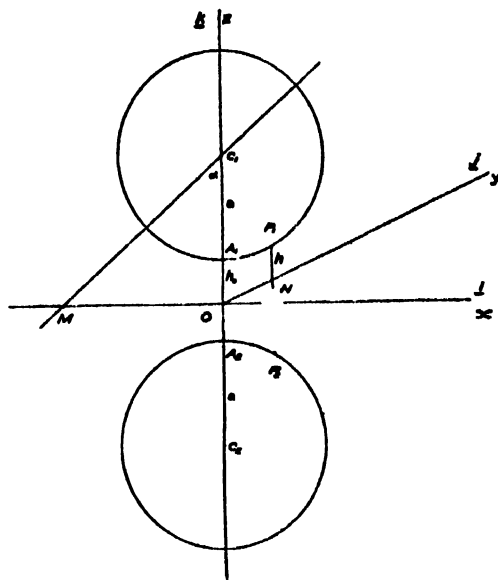


FIG. 2. Geometry of the 2-sphere system considered.

of interest, however; for example, the instrument could be used as a viscometer, for which purpose it would have the advantage of permitting wide variations in the working conditions, including the rate of shear; the relations obtained here provide a basis for the calibration of the instrument for that purpose.³ It was partly to this end that the determination of the torque was made. Also there is the point of mathematical interest that the investigation leads to an exact solution of the Reynolds equation, which is very seldom possible; it has in fact been stated⁴ that the equation cannot be solved in finite terms if the boundaries of the flow have curvature in two directions.

1. THE PRESSURE EQUATION (REYNOLDS EQUATION)

In Fig. 2, C_1 , C_2 are the centers of two spheres of radius a ; the first sphere rotates with angular velocity ω about the line MC_1 (which in the four-ball machine is vertical) and the second sphere is at rest; the inclination of MC_1 to C_1C_2 is taken as an arbitrary angle α whose value in the final application will be deduced from the geometry of the instrument.

³ It would be necessary to make a series of calibrations with liquids of widely differing viscosities to allow for possible changes in the dynamical pattern of the flow.

⁴ See, for example, R. O. Boswell, *Film Lubrication* (Longmans Green and Company, New York, 1928).

The coordinate system $Oxyz$ is chosen so that (a) O is the mid-point of C_1C_2 , (b) Oz coincides with the line of centers C_1C_2 , and (c) the axis MC_1 lies in the plane xz (in the figure, M lies on Ox).

Let p be the pressure and $\mathbf{v}$ the fluid velocity at any point; let the thickness of the film at any section be $2h$, and let the minimum thickness be $2h_0$. Thus in the figure $NP_1 = NP_2 = h$, $OA_1 = OA_2 = h_0$ and P_1P_2 have coordinates (x, y, h) and $(x, y, -h)$, respectively.

By assumption A , p is a function of x and y only, and the whole treatment of the problem depends on the partial differential equation in x and y satisfied by p . The derivation of this equation is a standard piece of analysis, but as we shall need in a later section some of the intermediate equations we may conveniently indicate here the main steps in the process.

If the fluid velocity $\mathbf{v}$ has components u, v, w , then the assumption B permits us to reduce the general equations of motion to

$$\frac{\partial p}{\partial x} = \mu \frac{\partial^2 u}{\partial z^2}, \quad \frac{\partial p}{\partial y} = \mu \frac{\partial^2 v}{\partial z^2}, \quad \frac{\partial p}{\partial z} = 0. \quad (1)$$

Let the velocity of the general point P_1 on the surface of the moving sphere have components U, V, W . Every point on the second sphere has zero velocity, so we can write the boundary conditions for these equations

$$\begin{aligned} \text{at } z = h \quad u = U \quad v = V \quad w = W; \\ \text{at } z = -h \quad u = 0 \quad v = 0 \quad w = 0. \end{aligned}$$

Hence by integration with respect to z

$$\begin{aligned} u &= \frac{z+h}{2h} U + \frac{z^2-h^2}{2\mu} \frac{\partial p}{\partial x}, \\ v &= \frac{z+h}{2h} V + \frac{z^2-h^2}{2\mu} \frac{\partial p}{\partial y}. \end{aligned} \quad (2)$$

Now the flow must satisfy the continuity equation $\text{div } \mathbf{v} = 0$

$$\text{i.e.,} \quad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \quad (3)$$

If we substitute for u and v the values given by (2) and then integrate the resulting equation

for w between the limits $-h$ and h so that

$$\int_{-h}^h \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) dz = - \int_{-h}^h \frac{\partial w}{\partial z} dz = -W,$$

we shall have eliminated the velocities u, v, w and derived an equation for p alone. We obtain

$$\begin{aligned} \frac{\partial}{\partial x} \left(h^3 \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left(h^3 \frac{\partial p}{\partial y} \right) \\ = \frac{3}{2} \mu W + \frac{3}{2} \mu h \left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} \right), \end{aligned} \quad (4)$$

which is the Reynolds equation, in rather more general form than is usually quoted.

So far the argument has been quite general and Eq. (4) holds for any system of the type we are considering; we must now obtain the analytical expressions for h, U, V, W which relate to our particular problem.

The expression for the film-thickness h is easily obtained, for if (x, y, h) is any point on the upper (moving) sphere, the distance from the center of the sphere is a , and so

$$x^2 + y^2 + (a + h_0 - h)^2 = a^2, \quad (5)$$

which defines h .

It is convenient to use the variable

$$a + h_0 - h = Z. \quad (5a)$$

Then

$$x^2 + y^2 + Z^2 = a^2, \quad (5b)$$

and the derivatives of h , which we shall need, are

$$\frac{\partial h}{\partial x} = - \frac{\partial Z}{\partial x} = \frac{x}{Z}, \quad (5c)$$

$$\frac{\partial h}{\partial y} = - \frac{\partial Z}{\partial y} = \frac{y}{Z}. \quad (5d)$$

The velocities U, V, W are most readily derived by a vector method. Let $\mathbf{i}, \mathbf{j}, \mathbf{k}$ be the usual unit vectors along the coordinate axes; then (Fig. 2) the vector angular velocity of the sphere is

$$\boldsymbol{\omega} = \omega(\sin \alpha \mathbf{i} + \cos \alpha \mathbf{k}),$$

and if $\mathbf{r}$ is the vector C_1P_1 the velocity of P_1 is $\boldsymbol{\omega} \wedge \mathbf{r}$. Now

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} - Z\mathbf{k},$$

so

$$U\mathbf{i} + V\mathbf{j} + W\mathbf{k} = \omega(\sin \alpha \mathbf{i} + \cos \alpha \mathbf{k}) \wedge (x\mathbf{i} + y\mathbf{j} - Z\mathbf{k}),$$

and equating the coefficients of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ on the two sides of this equation gives the required expressions:

$$\begin{aligned} U &= -\omega y \cos \alpha, \\ V &= \omega(x \cos \alpha + Z \sin \alpha), \\ W &= \omega y \sin \alpha, \end{aligned} \quad (6)$$

from which we have

$$\frac{\partial U}{\partial x} = 0, \quad \frac{\partial V}{\partial y} = -\frac{\omega y}{Z} \sin \alpha.$$

Substitution of these values in (4) gives as the fundamental equation for the problem

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{3}{2h} \left(x \frac{\partial p}{\partial x} + y \frac{\partial p}{\partial y} \right) - \frac{3\mu\omega \sin \alpha}{2h^3} \left(1 - \frac{h}{Z} \right) y = 0. \quad (7)$$

We can at once simplify the form of this equation by changing from Cartesian coordinates x , y to polars r , θ defined by

$$x = r \cos \theta, \quad y = r \sin \theta.$$

The transformed equation is

$$\begin{aligned} \frac{\partial^2 p}{\partial r^2} + \left(\frac{1}{r} + \frac{3r}{Zh} \right) \frac{\partial p}{\partial r} + \frac{1}{r^2} \frac{\partial^2 p}{\partial \theta^2} \\ = \frac{3\mu\omega \sin \alpha}{2h^3} \left(1 - \frac{h}{Z} \right) r \sin \theta, \end{aligned} \quad (8)$$

where

$$h = a + h_0 - Z, \quad Z^2 = a^2 - r^2,$$

and a most important point is that h and Z are functions of r only.

In this as in most problems it is of great advantage to work in terms of dimensionless variables; a little algebraic manipulation shows that if we define the dimensionless quantities ρ , Π , s_0 by the relations

$$\rho = r/(2ah_0)^{1/2}, \quad \Pi = p/\frac{3}{2}\mu\omega \sin \alpha (2a/h_0)^{1/2}, \quad s_0 = h_0/a,$$

the equation can be written

$$\frac{\partial^2 \Pi}{\partial \rho^2} + L \frac{\partial \Pi}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 \Pi}{\partial \theta^2} = M \sin \theta,$$

where L , M are functions of ρ only, given by

$$L = -\frac{1}{\rho} + \frac{3\rho}{(1-2s_0\rho^2)^{1/2}}(1+s_0-2s_0\rho^2), \quad (9a)$$

$$M = \rho \left[\frac{1}{2} \frac{1+s_0+(1-2s_0\rho^2)^{1/2}}{1+s_0+\rho^2} \right]^2 \times [2-(1+s_0)(1-2s_0\rho^2)^{1/2}]. \quad (9b)$$

If finally we regard $s_0 = h_0/a$ as a small quantity, and in (9a, b) neglect s_0 in comparison with unity, we obtain the simpler forms

$$L = -\frac{1}{\rho} + \frac{6\rho}{1+\rho^2}, \quad (9c)$$

$$M = \frac{\rho}{(1+\rho^2)^3}. \quad (9d)$$

In this last case we can write the equation in the form

$$\begin{aligned} \frac{\partial^2 \Pi}{\partial \rho^2} + \left(\frac{1}{\rho} + \frac{6\rho}{1+\rho^2} \right) \frac{\partial \Pi}{\partial \rho} \\ + \frac{1}{\rho^2} \frac{\partial^2 \Pi}{\partial \theta^2} = \frac{\rho}{(1+\rho^2)^3} \sin \theta, \end{aligned} \quad (10)$$

which is remarkable in that it is completely free from all the parameters of the problem: all these have been absorbed into the dimensionless variables Π , ρ . This single equation therefore covers all possible values of a , μ , α , ω ; but in dealing with this form we must remember that the original Eq. (9) has a singularity at $r = a$ (where $Z = 0$) so the range of ρ must be restricted to correspond to a range of r stopping well short of this point. This is, of course, quite in accordance with physical ideas; for at all points on the circle $r = a$ the tangent planes to the two spheres are parallel to the z axis (to put it less formally, the surfaces are diverging infinitely rapidly) and well before this the film will have become thicker than the basic assumptions will permit.

2. BOUNDARY CONDITIONS

Before we can make any use of the pressure equation we must ascribe to p a set of boundary conditions, and it is here that the real difficulty of the problem is met. We are dealing with an elliptic equation for p in the two independent

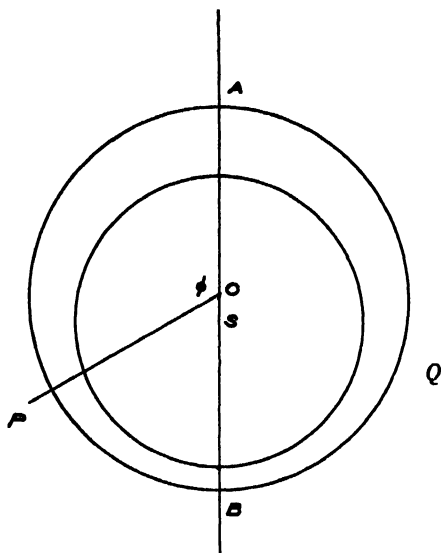


FIG. 3. Geometry of the complete journal bearing. S , O are the centers of the journal and the bearing, respectively.

variables x , y (or the equivalent variables r , θ) and general theory tells us that to obtain a unique solution we need to know the values of p at all points of a closed curve in the xy plane. The possibility of finding a solution in reasonably simple form is intimately bound up with the choice of this curve and of the set of values of p along the curve.

The difficulty arises because the film has no definite boundaries defined *a priori* by the geometry of the apparatus, but merges gradually into the main body of the fluid where the assumptions **A** and **B** are not valid. This is not to say that the problem is not defined physically: strictly the boundaries are the free surface of the liquid, the inner surface of the cup, and the four spherical surfaces; and the problem is defined by the obvious conditions of pressure and velocity at these surfaces. But in view of the small chance of finding a strict solution we must look for an alternative set of conditions for our simplified system, relying on general physical principles in making our choice.

An obvious condition is that in the main body of the liquid, and in fact everywhere except in the immediate vicinity of the point of closest approach of the two spheres, the pressure must be practically atmospheric; if, therefore, we choose the atmospheric value as the zero of pressure the

solution for p must reduce to very small (numerical) values everywhere except in this region, that is, for all except small values of r . This can be idealized to the statement

$$p=0 \quad \text{for } r \geq r_0,$$

where r_0 is some value considerably less than a , and suggests that for the boundary condition we might take

$$p=0 \quad \text{for } r=r_0, \quad 0 \leq \theta \leq 2\pi,$$

i.e., the vanishing of p on a circle of radius r_0 . This however is not a suitable condition; for the equation for p has what we may call "skew symmetry" about the origin: if in (7) we replace x , y by $-x$, $-y$ we merely change the sign of the right side. We get the same effect by replacing θ by $\theta + \pi$ in (8) or (9). If therefore the set of boundary values has the same property, the whole solution will itself have this property; that is, in polars, if in the region $0 < \theta < \pi$ the solution is $p=p(r, \theta)$ then in the region $\pi < \theta < 2\pi$ the solution will be $p=p(r, \theta + \pi) = -p(r, \theta)$. Since the suggested set of values has in fact this property, the use of this condition will lead to a solution in which positive and negative values of equal magnitude appear in pairs; the immediate conclusion, that the resultant load supported by the film is zero, is unacceptable, and not less so is the implication that in half the film the liquid is in a state of negative pressure.

At this stage we can profit from a glance at the theory of the complete journal bearing, the geometry of which is shown in Fig. 3. The classical treatment of Sommerfeld⁵ deals with a bearing of infinite length (so that no flow in the axial direction need be considered) and effectively solves the Reynolds' equation corresponding to the flow in the annular space between the two cylindrical surfaces, with the condition that the pressure and its derivatives are continuous and periodic functions of the angular coordinate ϕ ; the striking feature of the solution is that it gives positive pressures in the region APB ($0 < \phi < \pi$) and equal negative pressures in the region AQB ($\pi < \phi < 2\pi$). This prediction of negative pressures has led to much criticism of Sommerfeld's work; and it seems to be generally agreed that his premises depart seriously from the truth in that

⁵ A. Sommerfeld, *Zeits. f. math. Physik* 50, 97-155 (1904).

they make no provision for the replacement of the lubricant which is forced out of the ends of any real bearing, however long; in practice a bearing is connected to a source of the lubricant so that the leakage is continuously replaced. A rigorous treatment incorporating this feature would require a strict solution of the Reynolds' equation taking into account both radial and axial flow, and this seems to present insuperable mathematical difficulties; but an approximate solution developed by Morgan and Muskat⁶ suggests that the introduction of a source does in fact suppress the negative pressures. A less sound but much simpler treatment has been given by Gümbel;⁷ this writer, having assumed the connection to a source at atmospheric pressure, say at the point of maximum film-thickness $\phi=0$, then neglects the axial flow and suggests three possible empirical conditions to determine p :

I. The pressure falls to zero at the section of minimum thickness and remains zero until the position of the source is reached (traveling round the bearing in the direction of rotation of the axle), i.e., $p=0$ when $\phi=\pi$, and for $\pi \leq \phi \leq 2\pi$.

II. The pressure has its maximum value at the section of minimum thickness and is zero between that point and the position of the source, i.e., $dp/d\phi=0$ when $\phi=\pi$, and $p=0$ for $\pi < \phi \leq 2\pi$.

III. p and $dp/d\phi$ vanish simultaneously (the position at which this occurs being determined by the two conditions) and $p=0$ between this point and the position of the source.

What Gümbel assumes is in effect that the lubricating film breaks from the moving surface either at or just beyond the narrowest section; the first two suggested conditions imply a discontinuity in either p or $dp/d\phi$; the third is a little less drastic but no less arbitrary in essence. Nevertheless Gümbel, using these as working hypotheses, finds that the numerical solutions corresponding to the different possibilities do not differ seriously amongst themselves and are in reasonable agreement with experimental observations. He therefore concludes that his treatment, if not well founded logically, does at least give useful numerical results. A further application of what is effectively Gümbel's third condi-

tion has been made by Peppler⁸ in a study of the lubrication between the outer surfaces of a pair of cylinders—e.g., a pair of rollers—and he too finds good agreement between theory and observation.

If we compare the two problems—the flow between the spheres and that in the cylindrical bearing—we see that they have one feature in common, namely that they both deal with the flow in a channel of which a large part is divergent in the direction of flow; and as it is a general result of the classical theory of viscous flow that negative pressures are associated with divergent flow, we have here the fundamental reason for our difficulty. The essential feature of Morgan and Muskat's solution is the introduction of the source, and clearly the conditions of the present problem do not permit this; hence we are led to try the application of Gümbel's empirical treatment. Here the main flow is along the y axis and we may suppose that the film breaks from the moving surface along the x axis (since this is the position of narrowest section); in polars, along the lines $\theta=\pi$, $\theta=2\pi$. Further, relying upon Gümbel's results, we may choose from the conditions I, II, III the one which gives the simplest solution, and we find that this is condition I. We therefore take the boundary conditions as—

$$\begin{aligned} p &= 0 \quad \text{for } \theta = \pi, \quad \theta = 2\pi; \quad 0 \leq r \leq r_0, \\ p &= 0 \quad \text{for } r = r_0, \quad \pi \leq \theta \leq 2\pi, \end{aligned} \quad (11a)$$

where r_0 is yet to be determined. Thus we investigate the solution of the pressure equation over a semi-circular region of the xy plane, and effectively neglect any contribution to the pressure from the flow in the diverging region.

3. SOLUTION OF THE PRESSURE EQUATION

The actual solution of the Reynolds equation (9) with the boundary conditions (11a) is easily seen to be

$$\Pi = g \sin \theta, \quad (12)$$

where g is a function of ρ only defined by the differential equation

$$\frac{d^2 g}{d\rho^2} + L \frac{dg}{d\rho} - \frac{1}{\rho^2} g = M, \quad (13)$$

with the boundary conditions

$$g = 0 \quad \text{when } \rho = 0, \quad \rho = \rho_0.$$

⁶ F. Morgan and M. Muskat, *J. App. Phys.* **9**, 393 (1938).

⁷ L. Gümbel, *Zeits. f. tech. Physik* **3**, 94 (March, 1922).

⁸ K. O. Peppler, *V.D.I. Forschungs.* 391 (1938).

TABLE I. Solutions for g .^a

ρ	$G_1 \times 10^3$	$G_2 \times 10^3$	$g \times 10^3$	$G_1 \times 10^3$	$G_2 \times 10^3$	$g \times 10^3$
0	0	0	0	0	0	0
0.1	0	10	-10	0	10	-10
2	1	19	-18	1	19	-18
3	3	28	-24	3	28	-25
4	6	36	-29	6	36	-30
5	11	43	-31	11	43	-32
6	16	48	-31	16	48	-32
7	22	53	-30	22	53	-31
8	27	57	-29	27	57	-30
9	32	61	-28	33	61	-28
1.0	38	63	-24	38	63	-25
1.1	43	66	-23	43	66	-23
1.2	47	68	-20	48	68	-20
1.3	51	70	-18	52	70	-18
1.4	55	71	-15	56	71	-15
1.5	58	73	-14	59	73	-14
1.6	60	74	-13	62	74	-12
1.7	63	75	-10	64	75	-11
1.8	65	77	-9	67	76	-9
1.9	67	77	-8	69	77	-8
2.0	69	78	-7	71	77	-7
2.0	69	78	-7	71	77	-7
2.2	72	80	-6	74	80	-6
2.4	75	82	-5	76	82	-6
2.6	77	83	-4	79	83	-4
2.8	79	84	-3	81	84	-3
3.0	80	85	-3	82	85	-3
3.2	82	87	-2	85	87	-2
3.6	84	88	-2	86	88	-2
3.8	85	89	-2	88	89	-1
4.0	86	90	-2	89	90	-1
4.2	87	91	-1	90	91	-1
4.4	88	91	-1	90	91	-1
4.6	89	92	-1	91	92	-1
4.8	90	92	0	92	92	0
5.0	91	93	0	93	93	0

^a g is defined by $g = G_1 + \lambda G_2$ with λ chosen so that $g = 0$ when $\rho = 5.0$. The three-decimal values given above have been obtained by rounding off the original four-decimal values.

For this function satisfies both the partial differential equation and the boundary conditions, and is therefore a possible solution. The general uniqueness theorem then tells us that this is the only solution.

All that remains now is to find the solution of Eq. (13), and although there seems to be no simple closed analytical form for the function $g(\rho)$ a numerical solution is very easily constructed.

This function can be expressed in the form

$$g = G_1 + \lambda G_2,$$

where G_1 is any solution of the equation

$$G_1'' + LG_1' - \frac{1}{\rho^2}G_1 = M \quad (13a)$$

(dashes denoting differentiation with respect to ρ),

and G_2 is a suitably chosen complementary function, that is a solution of

$$G_2'' + LG_2' - \frac{1}{\rho^2}G_2 = 0, \quad (13b)$$

and λ is a numerical constant to be chosen later. With the forms of L , M given by (9a) and (9b) it is readily seen that:

(a) The only solutions for G_1 and G_2 which are finite at $\rho = 0$ must vanish at this point; near $\rho = 0$ they have the forms

$$G_1 = \frac{1}{2}\rho^2[1 - \frac{1}{4}\rho^2 + \dots],$$

$$G_2 = \rho[1 - \frac{3}{4}\rho^2 + \dots].$$

(b) Neither G_1 nor G_2 can vanish for any other value of ρ . Therefore if $g = G_1 + \lambda G_2$ we have at once $g = 0$ for $\rho = 0$, and we can find λ so that $g = 0$ for $\rho = \rho_0$ [in fact, $\lambda = -G_1(\rho_0)/G_2(\rho_0)$]. The problem is now reduced to the integration of Eqs. (13a), (13b) with the conditions

$$G_1 = 0, \quad G_1' = 0 \quad \text{at} \quad \rho = 0,$$

$$G_2 = 0, \quad G_2' = 1 \quad \text{at} \quad \rho = 0,$$

which can be carried out by standard technique.

Since there is a choice of two forms for the functions L , M two distinct integrations were made, one with the approximate forms of (9c), (9d) and one with the "exact" forms of (9a), (9b) taking $h_0/a = 1/200$ —this should rank as rather a large value, as we shall see later. These numerical results are given in Table I, and it is at once clear that the error of the approximation is negligible. But the most important property of the functions G_1 and G_2 revealed by the numerical solutions is that they quickly attain virtually constant values; more particularly, their values remain sensibly constant for $\rho \geq 5$. It follows that for any given value of λ the function $g = G_1 + \lambda G_2$ is itself sensibly constant for $\rho \geq 5$. Therefore if we take for ρ_0 any value not less than 5, we shall have, with $\lambda = -G_1(\rho_0)/G_2(\rho_0)$, $g = 0$ for $\rho = \rho_0$. In other words, although the quantity ρ_0 is not formally defined its choice need cause no difficulty. The value chosen for the solution of Table I was $\rho_0 = 5.0$.

A little reflection shows that this lack of sensitivity to the exact position of the boundary is to be expected on physical grounds. For in the immediate vicinity of the region of closest ap-

proach of the two spheres—which is the only region we consider—the pressure must fall off very rapidly with increasing distance from the actual point of closest approach, and therefore so long as we ensure that the boundary is in the region where the pressure has fallen to a small value, the exact position cannot have any great effect on the solution.

4. TOTAL LOAD SUPPORTED BY THE FILM

Since the pressure p does not vary through the thickness of the film the total load supported is simply the integral of p over the xy plane.

Thus,

$$\begin{aligned}\text{load } P &= \int \int p dx dy = \int_r^{2r} \int_0^{\pi_0} p r dr d\theta \\ &= \frac{1}{2} \mu \omega \sin \alpha \left(\frac{2a}{h_0} \right)^{\frac{1}{2}} \cdot 2a h_0 \int_r^{2r} \int_0^{\pi_0} \Pi \rho d\rho d\theta \\ &= 12 K a^2 \mu \omega \sin \alpha (2a/h_0)^{\frac{1}{2}} \text{ since } \Pi = g \sin \theta,\end{aligned}$$

where

$$K = - \int_0^{\pi_0} \rho g(\rho) d\rho.$$

From the solutions given in Table I we find

for $s_0 = 1/200$, $K = 0.057$,

$s_0 = 0$, $K = 0.053$ (the approximate case).

Clearly there is no point in keeping many decimals in a quantity such as K , so we may take as a round value $K = 1/20$ and write the final result as

$$P = \frac{1}{2} a^2 \mu \omega \sin \alpha (2a/h_0)^{\frac{1}{2}}. \quad (14)$$

It is from this equation alone that we derive our criterion for the range of validity of the classical theory; but it is convenient to postpone this until we have obtained the corresponding relation for the torque exerted by the viscous-frictional forces, and have applied both results to the particular orientation of spheres found in the four-ball instrument. These topics are dealt with in the next two sections.

5. TORQUE ON THE STATIONARY SPHERE

As was explained in the Introduction it was considered desirable to obtain the expression for the torque exerted on the stationary sphere in the

four-ball apparatus, and this section gives a derivation of the corresponding relation for the pair of spheres considered in the preceding sections. It cannot be claimed that the result obtained here gives anything more than an indication of the leading term of the required relation. For on the one hand we can take into account only the conditions in the region where the flow is converging, since we know nothing of the conditions in the diverging region; and on the other, the unknown dimension r_0 (of which the expression for the load was found to be effectively independent) appears explicitly in the final expression. Nevertheless the result should show the essential features of the relation.

The problem is to find the viscous-frictional torque on a curved surface immersed in a body of liquid whose motion is essentially three-dimensional; this is necessarily rather complicated, and after various attempts to construct an approximate solution it was decided that the only satisfactory method was to use the fundamental tensor-relations for the shear-stress in the liquid. Only the final expression for the torque is needed here, and it would serve no useful purpose to give at this stage the rather heavy algebra of the derivation; but as no explicit treatment of the general case was found in the standard textbooks the analysis is given in an Appendix.

If Γ is the torque about the axis of rotation exerted by the fluid on the stationary sphere, we find the approximate relation

$$\Gamma / \frac{1}{2} \pi a^3 \mu \omega = \frac{1}{2} + \left[\log_e \left(\frac{a}{h_0} \right) - 1 \right] \sin^2 \alpha,$$

and if we put $\log_e (a/h_0) = 2.3 \log_{10} (a/h_0)$,

$$\begin{aligned}\Gamma / \frac{1}{2} \pi a^3 \mu \omega &= \left(\frac{1}{2} - \sin^2 \alpha \right) \\ &\quad + 2.3 \sin^2 \alpha \log_{10} \left(\frac{a}{h_0} \right). \quad (15)\end{aligned}$$

6. APPLICATION TO THE FOUR-BALL INSTRUMENT

To derive from the foregoing results the expressions for the load and torque for the four-ball instrument we have only to determine the angle α between the axis of the rotating sphere and the line joining the center of this sphere to the center of any one of the fixed spheres.

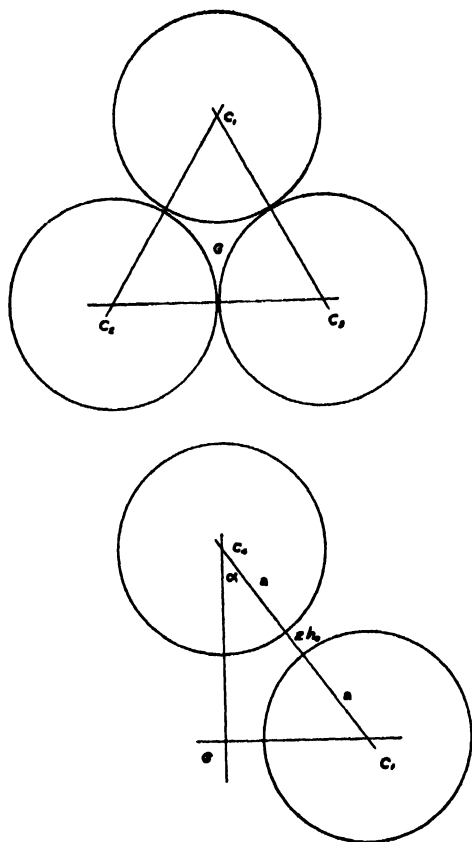


FIG. 4. Geometry of the 4-ball instrument.

The geometry of the system is shown in Fig. 4. It is clear that the vertical axis of rotation must pass through the centroid G of the triangle formed by the centers C_1, C_2, C_3 of the fixed spheres.

Then

$$C_1G = \frac{2\sqrt{3}}{3}a,$$

and so, from the triangle C_1GC_2 ,

$$\sin \alpha = GC_1/C_1C_2 = \frac{2\sqrt{3}}{3}a / 2(a+h_0) = \left(\frac{1}{2}\right)^{\frac{1}{2}} \cdot \frac{a}{a+h_0}.$$

Since $h_0 \ll a$, we may take

$$\sin \alpha = \left(\frac{1}{2}\right)^{\frac{1}{2}}, \quad \cos \alpha = \left(\frac{1}{2}\right)^{\frac{1}{2}} \quad (\alpha = 35^\circ 16'). \quad (16)$$

The load P determined in §4 acts along the line of centers C_1C_2 ; by symmetry therefore the total load supported by the three spheres has zero horizontal component and the vertical component is $3P \cos \alpha$. Thus the total load on the axis

is, by (14) and (16),

$$W = 3P \cos \alpha = -\frac{6}{5}a^2\mu\omega \left(\frac{2a}{h_0}\right)^{\frac{1}{2}}. \quad (17)$$

The total torque on the axis is, by (15) and (16),

$$T = 3\Gamma = \frac{3}{2}\pi a^3\mu\omega \left[-\frac{2}{15} + 0.77 \log_{10} \left(\frac{a}{h_0} \right) \right]. \quad (18)$$

If now we eliminate a/h_0 from these relations, we get the relation between T and W

$$T / \frac{3}{2}\pi a^3\mu\omega = -\frac{2}{15} + 1.54 \log_{10} \left[W / \frac{6\sqrt{2}}{5}a^2\mu\omega \right], \quad (19)$$

showing that if experimental values of T and W are found and $T/a^3\mu\omega$ is plotted against $\log (W/a^2\mu\omega)$, then the points should lie on a straight line.

The sizes of the quantities involved are best shown by a numerical example. In applying the relations we must remember that all quantities are presumed to be given in absolute units, so that if a is measured in centimeters, ω in radians per sec., and μ in c.g.s. units (poises) then W will be given in dynes and T in dyne-centimeters.

The usual size of instrument corresponds to spheres of 1-cm diameter, so $a = \frac{1}{2}$; for a rather viscous lubricant we may take $\mu = 10$ c.g.s. units at moderate temperatures. Let us take $\omega = N$ r.p.m. $= 2\pi N/60$ radians per sec. and $h_0/a = 10^{-n}$ (so that $h_0 = \frac{1}{2} \cdot 10^{-n}$ cm).

A normal speed of rotation is 1500 r.p.m. and with this value we find (rounding off the numerical coefficients)

$$W = 500 \times 10^{n/2} \text{ dynes} = 0.51 \times 10^{n/2} \text{ g wt.},$$

$$T = 720n - 125 \text{ dyne cm} = 0.74n - 0.13 \text{ g cm.}$$

By taking different values for n we obtain a series of related values of load and torque corresponding to a series of film-thicknesses h_0 ; the results for the range $2 \leq n \leq 6$ (i.e., for film-thicknesses ranging from 5×10^{-8} to 5×10^{-7} cm) are given in Fig. 5.

It is now clear that no great loads can be supported by the apparatus if the conditions of film-lubrication are to be fulfilled. There is a lower limit to the thickness of the film, fixed by the size of the irregularities of the surfaces, for

which a value of the order of 10^{-4} cm has been suggested. With this value and even the rather viscous oil considered above, the load W cannot greatly exceed 50 g wt.; even if the speed were raised to 3000 r.p.m., the permissible load would rise only to about 100 g wt. Evidently a load of the order of a kilogram represents a virtual impossibility for an instrument of this size under conditions of film-lubrication, unless it can be established that this regime can be maintained with film-thicknesses of the order of 10^{-6} cm. It is quite certain that under the usual conditions of operation—with loads of 20 to 1000 kg—there can be no hydrodynamical lubrication; this is of course only what was expected, but the investigation has provided the criterion from which we can deduce the range of conditions in which the behavior of the apparatus obeys the laws of hydrodynamical lubrication.

APPENDIX. TORQUE ON THE STATIONARY SPHERE

The stress at any point in a viscous liquid has six independent components which form a symmetrical tensor of the second rank. This tensor is

$$\Psi = -pI + \mu T,$$

where I is the unit tensor and T is the shear-stress tensor, and these can be conveniently written in "dyadic" form

$$I = \mathbf{i}\mathbf{i} + \mathbf{j}\mathbf{j} + \mathbf{k}\mathbf{k},$$

$$T = 2 \sum \frac{\partial u}{\partial x} \mathbf{i}\mathbf{i} + \sum \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right) (\mathbf{j}\mathbf{k} + \mathbf{k}\mathbf{j}),$$

where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are as before the unit vectors along the coordinate axes.⁹

The force on a surface of unit area whose normal has the direction of the unit vector $\mathbf{n}$ is itself a vector and is given by the scalar product of $\mathbf{n}$ with the stress tensor in the form

$$\mathbf{F}_n = \Psi \cdot \mathbf{n} = -p\mathbf{n} + \mu T \cdot \mathbf{n}.$$

Now if a force $\mathbf{F}$ is applied at a point P whose position vector with respect to a fixed point M is $\mathbf{r}$, the moment of $\mathbf{F}$ about M is $\mathbf{r} \wedge \mathbf{F}$; and if through M there passes a fixed axis whose direction is the unit vector $\mathbf{t}$ then the torque about

⁹ C. E. Weatherburn, *Advanced Vector Analysis* (G. Bell and Sons, London, 1928), p. 149.

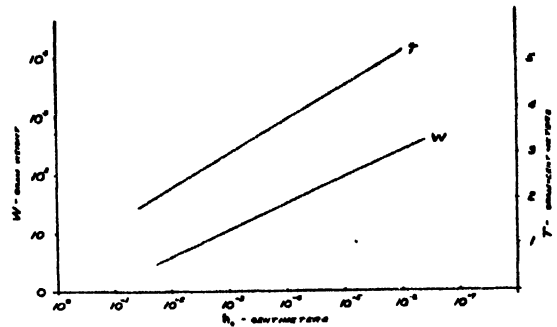


FIG. 5. Numerical values for load (W) and torque (T) for varying minimum film-thickness (h_0).

this axis is $\mathbf{t} \cdot \mathbf{r} \wedge \mathbf{F}$. Thus if for $\mathbf{F}$ we take the hydrodynamical force $\mathbf{F}_n$, we have the general result that the torque γ about the axis $\mathbf{t}$ of the force exerted by the liquid on the unit area at the point $\mathbf{r}$ is

$$\gamma = \mathbf{t} \cdot \mathbf{r} \wedge \mathbf{F}_n = \mathbf{t} \cdot \mathbf{r} \wedge (-p\mathbf{n} + \mu T \cdot \mathbf{n}),$$

where it is understood that the pressure p and the stress-tensor T are evaluated at the point $\mathbf{r}$.

The resultant torque Γ of the force exerted on the surface S in the liquid is then simply the integral of γ

$$\text{i.e.,} \quad \Gamma = \int \gamma dS$$

taken over the surface S .

This result is perfectly general; in any particular case we have only to substitute the appropriate values for $\mathbf{t}$ (given by the position and direction of the axis), $\mathbf{r}$, $\mathbf{n}$ (given by the form and position of the surface), and p , T (given by the velocity distribution of the flow).

We are interested in the torque, about the axis of rotation of the moving sphere, of the force exerted by the liquid on the fixed sphere; we therefore choose for $\mathbf{t}$, $\mathbf{r}$, $\mathbf{n}$ the following values (Fig. 2).

MC_1 is the axis of rotation of the moving sphere and P_2 the general point $(x, y, -h)$ on the fixed sphere.

Then

$$\begin{aligned} \mathbf{t} &= \sin \alpha \mathbf{i} + \cos \alpha \mathbf{k}, \\ \mathbf{r} &= \mathbf{MP}_2 = [x + (a + h_0) \tan \alpha] \mathbf{i} + y \mathbf{j} - h \mathbf{k}. \end{aligned}$$

The surface being spherical, the normal at any point has the same direction as the radius drawn

to that point; therefore,

$$\mathbf{an} = \mathbf{C}_2 \mathbf{P}_2 = x\mathbf{i} + y\mathbf{j} + Z\mathbf{k}.$$

Evaluating the triple product we find

$$\gamma = (yF_z + ZF_y) \sin \alpha + (xF_y - yF_x) \cos \alpha + 2hF_y \sin \alpha,$$

where F_x, F_y, F_z are the components of $\mathbf{F}_n$.

In order to evaluate the tensor components, we need explicit forms for the pressure p and the velocity derivatives $\partial u/\partial x$, etc. The Eqs. (2) and the continuity Eq. (3) together determine these latter; all except $(\partial w/\partial x)$, $(\partial w/\partial y)$ are given immediately, and for these we have first to determine w by integration of $\partial w/\partial z$ with the condition $w=0$ when $z=-h$, and then to differentiate these with respect to x and y in turn. Putting $z=-h$ in these expressions then gives the values required. Since

$$\begin{aligned} \mathbf{aF}_n &= \mathbf{an} \cdot \Psi = -(x\mathbf{i} + y\mathbf{j} + Z\mathbf{k})p \\ &\quad + \mu(x\mathbf{i} + y\mathbf{j} + Z\mathbf{k}) \cdot \mathbf{T}, \\ &= F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}, \end{aligned}$$

we can now express the force components as functions of x and y and hence also the torque γ ; the algebra is straightforward but rather heavy, and the final result is

$$\begin{aligned} \frac{2Zh}{\mu\omega a} \gamma &= (x^2 + y^2) \cos^2 \alpha + 2(a + h_0)x \cos \alpha \sin \alpha \\ &\quad + \left[(a^2 + x^2) + \frac{3}{2} \left(1 - \frac{h}{Z} \right) y^2 + 2hZ \right] \sin^2 \alpha \quad (a) \\ &\quad - \frac{2h^2}{\mu\omega} \left(x \frac{\partial p}{\partial y} - y \frac{\partial p}{\partial x} \right) \cos \alpha \quad (b) \\ &\quad - \frac{h^2}{\mu\omega Z} \left[y \left(x \frac{\partial p}{\partial x} + y \frac{\partial p}{\partial y} \right) + 2Z^2 \frac{\partial p}{\partial y} \right] \sin \alpha \quad (b) \\ &\quad - \frac{4h^3}{\mu\omega} \frac{\partial p}{\partial y} \sin \alpha + \frac{4h^4 Z y}{a^2 \mu\omega} (\nabla^2 p) \sin \alpha \quad (A) \end{aligned}$$

(the indications "a" and "b" will be explained later).

We have now to integrate this expression over a suitable area of the spherical surface; if we introduce the plane polar coordinates r, θ as in §2 we have for the element of area the expression

$$dS = r dr d\theta / (1 - r^2/a^2)^{1/2},$$

and the limits for the integration are $\pi \leq \theta \leq 2\pi$, $0 \leq r \leq r_0$. Alternatively we can use a pair of angular coordinates θ, ϕ , where θ is defined as before and ϕ by the relation $r = a \sin \phi$. The limits are now $\pi \leq \theta \leq 2\pi$, $0 \leq \phi \leq \phi_0$, where $r_0 = a \sin \phi_0$, and the element of area is

$$dS = a^2 \sin \phi d\theta d\phi.$$

The first set of coordinates is more convenient for dealing with the terms in γ which involve p explicitly, and the second for the other terms. We have also the relations

$$Z = a + h_0 - h = (a^2 - r^2)^{1/2} = a \cos \phi,$$

$$r = \rho(2ah_0)^{1/2}$$

$$p = \frac{3}{2} \mu \omega \sin \alpha \left(\frac{2a}{h_0} \right)^{1/2} g(\rho) \sin \theta,$$

the function g being given numerically (Table I).

Examination of the terms of (A) shows that those marked (a) vanish identically on integration and that those marked (b) lead to expressions which are of higher order in h/a than the remainder, and which may therefore be neglected since this is a small quantity and we want only the leading term in the expression. Thus we have to integrate only the terms

$$\frac{x^2 + y^2}{hZ}, \quad \left(\frac{a^2 + \frac{1}{2}y^2}{hZ} + 2 \right) \sin^2 \alpha, \quad \frac{2h}{\mu\omega} \frac{\partial p}{\partial y} \sin \alpha.$$

The integration is straightforward and the result is

$$\begin{aligned} \Gamma / \frac{1}{2} \pi a^2 \mu \omega &= 1 - \cos \phi_0 - \log \cos \phi_0 \\ &\quad + \sin^2 \alpha [2(1 - \cos \phi_0) \\ &\quad + \frac{1}{2}(1 - \cos \phi_0 - \log \cos \phi_0) \\ &\quad + \log(1 + s_0 - \cos \phi_0) / s_0 \cos \phi_0] \\ &\quad + O(s_0 \log s_0) - 12 \sin^2 \alpha \int_0^{s_0} \rho^2 g d\rho, \end{aligned}$$

where all logarithms are to the base e .

The leading terms of this expression can be written

$$\Gamma/\frac{1}{2}\pi a^3\mu\omega = A + (\log a/h_0 + B) \sin^2 \alpha,$$

where

$$A = 1 - \cos \phi_0 - \log \cos \phi_0$$

$$B = 2(1 - \cos \phi_0) + \frac{1}{2}(1 - \cos \phi_0 - \log \cos \phi_0)$$

$$+ \log [(1 - \cos \phi_0)/\cos \phi_0] - 12 \int_0^{\phi_0} \rho^2 g d\rho.$$

From the values of $g(\rho)$ in Table I we find

$$-12 \int_0^{\phi_0} \rho^2 g d\rho = -12 \times (-0.098) = 1.2 \text{ approx.},$$

but we cannot hope to determine the values of the terms involving $\cos \phi_0$ with any precision, since we have no precise knowledge of the value of ϕ_0 . However we can obtain a rough estimate by the following argument. ϕ_0 is defined by $a \sin \phi_0 = r_0$, where r_0 measures the extent of the region to which the film-theory may be applied. This dimension is determined by the thickness h of the film, and there is the relation

$$h = a + h_0 - (a^2 - r^2)^{1/2} = a + h_0 - a \cos \phi.$$

If we can say that the theory can be applied for thicknesses up to a value h_m , then we can take for ϕ_0 the value given by

$$\cos \phi_0 = 1 - (h_m - h_0)/a.$$

We cannot say that this maximum permissible thickness has a definite value, since the value must depend upon the particular conditions prevailing; but we may not unreasonably suppose it to be considerably greater than the greatest value (10^{-2} cm) of h_0 considered here, so that ϕ_0 is effectively independent of h_0 , and $\cos \phi_0 = 1 - h_m/a$. If as a working hypothesis we assume that h_m is not less than about a millimeter, we have the figure 0.1 as an indication of the order of lower limit of h_m/a (since the radius a is of the order of unity) and 0.9 for the upper limit of $\cos \phi_0$.

There is also the relation $\sin \phi_0 = r_0/a$, and the analysis given in the preceding pages makes considerable use of the assumption that r_0 is considerably less than a (since the exact equation for the pressure has a singularity at $r=a$). We may suppose $r_0/a < \frac{1}{2}$, and hence it is suggested that $\cos \phi_0 > \frac{1}{2}(3)^{1/2} = 0.866$.

These arguments are very specious, but they suggest that the order of the constants A , B will be given by taking a value in the region of 0.9 for $\cos \phi_0$; if we tabulate the values for a few values of $\cos \phi_0$ we get

$\cos \phi_0$	A	B
0.95	0.1	-1.7
.90	0.2	-0.9
.85	0.3	-0.3
.80	0.4	+0.1

so we shall not be seriously in error if we choose the round values

$$A = \frac{1}{2}, B = -1.$$

Thus we have the result that the torque is a linear function of $\log a/h_0$ having approximately the form

$$\Gamma/\frac{1}{2}\pi a^3\mu\omega = \frac{1}{2} + \left[\log_a \left(\frac{a}{h_0} \right) - 1 \right] \sin^2 \alpha.$$

This indicates a very slow variation with a/h_0 and it is for this reason that the expression for γ was evaluated in full; it is difficult to ensure that a "leading term" type of argument applied early in the calculation does not neglect terms which are in fact significant.

ACKNOWLEDGMENTS

This investigation was made at the request of Dr. R. Schnurmann of the Manchester Oil Refinery Ltd. (England). The author wishes to express his appreciation of Dr. Schnurmann's personal interest in the work and of his valuable discussion and criticism, and of his company's permission to publish the work. Also the author acknowledges with gratitude the advice he has had from Professor D. R. Hartree, F.R.S.

The Analysis of Multicomponent Mixtures of Hydrocarbons in the Liquid Phase by Means of Infra-Red Absorption Spectroscopy

D. L. FRY, R. E. NUSBAUM

Research Laboratories Division, General Motors Corporation, Detroit, Michigan

AND

H. M. RANDALL

Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan

(Received October 17, 1945)

This is a general method of applying infra-red absorption spectroscopy to the analysis of multicomponent hydrocarbon mixtures in the liquid phase. The procedure is calibrated by measuring optical densities of synthetic standard samples. A constant thickness cell is used for both standard and unknown samples. Two procedures for converting optical densities to concentrations are described. Examples are given for four- and five-component mixtures. Data are presented to show the reproducibility of repeated measurements of optical density on the same sample, and results are given to show the agreement between the infra-red analyses and the known composition of synthetic mixtures.

INTRODUCTION

A GENERAL method of applying infra-red absorption spectroscopy to the analysis of multicomponent hydrocarbon mixtures in the liquid phase has been developed and has been applied to two particular mixtures, one containing four components and the other five components. It is an empirical method based upon analytical curves prepared from absorption measurements of pure samples of the components to be found in the unknowns and of binary mixtures of these components.

A number of papers on the application of infra-red absorption spectroscopy to analytical problems have been published. Most of these have been concerned with the analysis of binary mixtures. However, two papers have described methods for the analysis of more complicated mixtures. Brattain, Rasmussen, and Cravath¹ have described a method for analyzing multicomponent mixtures in the vapor phase. This method is most useful when the components are in the vapor phase under normal conditions of temperature and pressure, but can also be used with considerably more difficulty when the compounds are liquids under normal conditions. A method for analyzing a multicomponent mixture in the liquid phase has been described by Nielsen

and Smith² and the analysis of a three-component mixture has been discussed as an example. This method can be used only over concentration ranges for which deviations from Beer's law are negligible.

The procedures described in the present paper are used for multicomponent mixtures in the liquid phase and are applicable even though the concentrations of any of the components of the mixture vary from 0 to 100 percent.

The absorption of monochromatic radiation as it traverses an absorbing liquid is expressed by Beer's law.

$$acx = \log_{10} (I_0/I), \quad (1)$$

where I is the intensity of the radiation transmitted by x centimeters of liquid containing a concentration c of the absorbing materials; I_0 is the intensity of the radiation impinging on the liquid; and a is a constant of proportionality, known as the absorption coefficient. a is constant for any monochromatic radiation regardless of the values of c , x , or I_0 .

For a mixture of two or more non-associating components, all of which selectively absorb somewhat over the whole infra-red region, Beer's law takes the following form for monochromatic radiation:

$$(a_1c_1 + a_2c_2 + a_3c_3 + \cdots + a_nc_n)x = \log_{10} \frac{I_0}{I}, \quad (2)$$

¹ R. Robert Brattain, R. S. Rasmussen, and A. M. Cravath, *J. App. Phys.* 14, 418 (1943).

² J. Rud Nielsen and Don C. Smith, *Ind. Eng. Chem. Anal. Ed.* 15, 609 (1943).

where a is the absorption coefficient of material (1) and c_1 its concentration in the cell of thickness x ; $a_2, a_3 \dots a_n$ and $c_2, c_3 \dots c_n$ are the corresponding values for materials 2, 3 $\dots n$.

In actual practice, however, Eq. (1) and, therefore, Eq. (2) does not hold exactly. This is shown experimentally by the fact that a curve of $\log I_0/I$ as a function of concentration with a constant thickness cell is not a straight line. This results from the fact that no spectrometer produces a monochromatic image at the receiver. Instead, there is produced an image containing a wave-length band $\Delta\lambda$.

Equation (3) is obtained by combining Eqs. (1) and (2).

$$\log \frac{I_0}{I} = \log \frac{I_0}{I_1} + \log \frac{I_0}{I_2} + \dots + \log \frac{I_0}{I_n} \quad (3)$$

The terms

$$\log \frac{I_0}{I_1}, \log \frac{I_0}{I_2}, \dots \log \frac{I_0}{I_n}$$

are the values obtained for materials 1, 2 $\dots n$ at the chosen spectral position. In other words, the equation states that the optical density of a mixture is equal to the sum of the optical

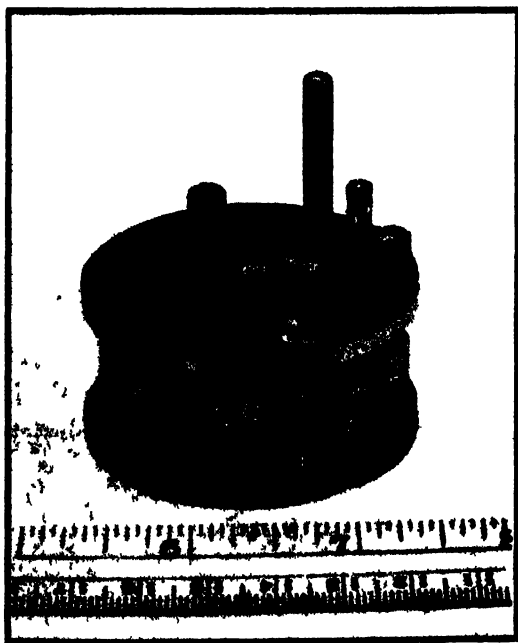


FIG. 1. An assembled cell used to obtain the infra-red absorption of liquids.

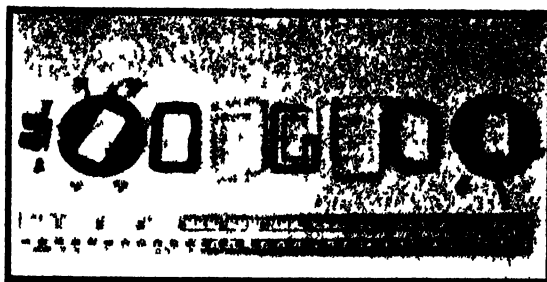


FIG. 2 Unassembled parts of a cell used in measuring the infra-red absorption of liquids.

densities of the individual components of the mixture Equation (3) does not contain the absorption coefficient and, therefore, is not subject to the same limitations as are (1) and (2).

DEVELOPMENT OF ANALYTICAL PROCEDURE

The problem in the work described here, as well as in the work of others, has been to develop a method of applying absorption spectroscopy taking into account these limitations of Beer's law. The fundamental problem of quantitative analysis in absorption spectroscopy is the determination of the $\log I_0/I$ values with sufficiently high accuracy, and the conversion of these values to percent concentration of the components in the mixture.

Absorption Cells

In the present work, optical densities are measured with a University of Michigan infra-red spectrophotometer³ equipped with a sodium chloride prism. Two cells are used. The background cell, consisting of a single salt plate $\frac{1}{4}$ " thick, is used to determine I_0 values. The absorption cell, shown in Fig. 1 and Fig. 2, is a modification of a cell described by Randall.⁴ The frame which holds the salt plates is designed particularly for use with the University of Michigan instrument.

Starting from the left, each piece in Fig. 2 goes on top of the previous one. The salt plates of the absorption cell are $\frac{1}{4}$ " thick. They are separated by a platinum shim, the thickness of

³ H. M. Randall and John Strong, Rev. Sci. Inst. 2, 585 (1931).

⁴ Robert A. Oetjen and H. M. Randall, unpublished thesis, University of Michigan, 1941.

⁵ H. M. Randall, Rev. Sci. Inst. 10, 195 (1939).

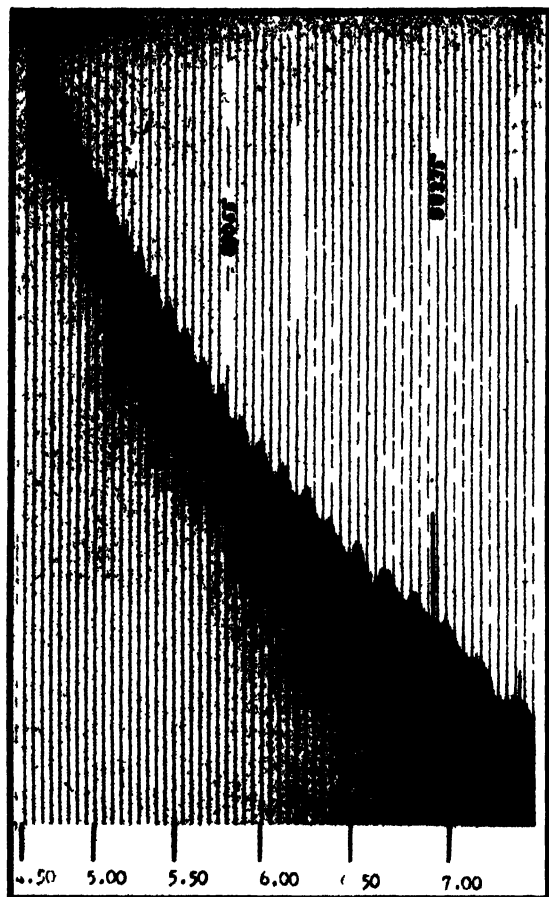


FIG. 3. A record of the interference fringe pattern obtained with a cell 0.152 mm thick.

which determines the thickness of the absorbing layer of sample. For the work described here the shim is 0.152 mm thick. The positions of the salt plates and the shim are determined by the dowel pins in the cell supports. Two rubber washers separate the salt from the stainless steel housing. After the cell is assembled, a mercury seal is established by introducing mercury into the trough in the lower plate through a hole which extends through the upper salt plate, rubber washer, and support. The liquid to be examined is introduced through one of the two holes on the side of the upper plate. Rubber pads are pressed over the holes to seal them without putting pressure on the contents of the cell. This method of sealing makes it possible to use the cell in an evacuated source box. The cell can be taken apart, cleaned, and reassembled in less than 10 minutes. However, it can be refilled in

less than two minutes by expelling the liquid with an air syringe and filling with an eye dropper.

In order to use the present method successfully, it is important that the thickness of the absorption cell remain constant. The two salt plates are ground flat with corundum in ethyl alcohol, and polished on a piece of white felt stretched over a piece of plate glass. Titanium oxide is used as the abrasive for polishing, and ethyl alcohol is used to wet the titanium oxide. The finished plates are examined with an optical flat and sodium light, and if they are not flat within four diffraction fringes (i.e., 0.0012 mm) the process is repeated. The plates are examined frequently to determine whether cleaning has caused deviation from flatness.

The cells are supported in a tube directly in front of the salt window at the entrance slit of the spectrometer. The notch in the frame of the cell locates it in a given fixed position in front of the slit. The cells are handled by means of a knurled rod screwed into the upper plate.

The thickness of the cell is determined by the interference fringe method as described by Smith and Miller.⁶ (See Fig. 3.) Repeat measurements made several months apart show the thickness to be 0.152 ± 0.0005 mm.

Spectral Positions

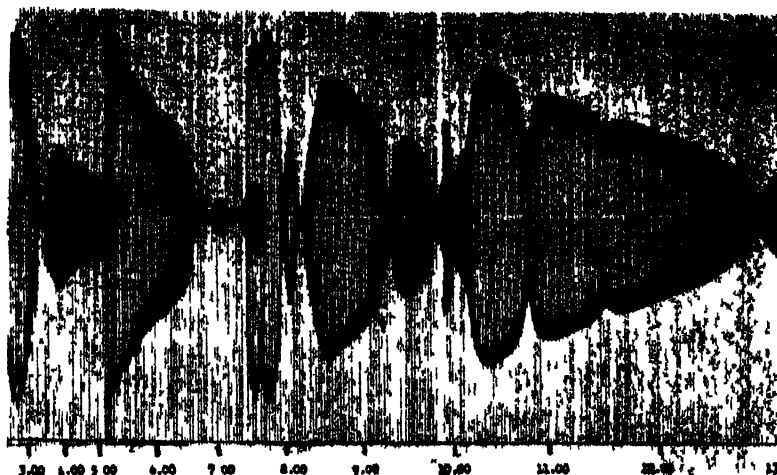
The approximate spectral positions at which optical densities are determined are selected from the records of the spectra of the pure components in the region from 3 to 13.5μ . (See Fig. 4.)

In order to obtain high resolution for these records the spectrograph is operated with narrow slits and slow rotation of the prism. The source of light is a Nernst glower. The light beam is interrupted by means of a shutter having a period of 3.8 seconds, and the output of the thermopile is amplified by a Firestone amplifier.⁷ The recording galvanometer is five meters from the camera drum. The spectra are recorded in small sections with constant slits, and an overlapping of these sections of spectrum occurs whenever the slit widths are changed. Table I gives the maximum theoretical resolution ob-

⁶ Don C. Smith and Elmer C. Miller, *J. Opt. Soc. Am.* **34**, 130 (1944).

⁷ F. A. Firestone, *Rev. Sci. Inst.* **3**, 163 (1932).

FIG. 4. The absorption spectrum of 2,2 dimethylbutane in the region 2.5 to 13.5 μ



tainable with the slit widths necessary to get half scale deflections with this amplifier.⁴ Actually the resolution realized with the instrument is somewhat less than the theoretical value.

In the analysis of an n component mixture it is necessary and sufficient to determine $\log I_0/I$ values of the mixtures at n different spectral positions. When a procedure is being developed for a given mixture, the spectra of the pure components are examined and n positions are selected in such a way that a different component contributes a maximum amount of absorption at each position. It is desirable but not necessary that at each wave-length position the absorption of one component be greater than that due to any other component; and further, that the component showing major absorption be different at each of the n wave-length positions. The spectral positions which yield most accurate analyses are those at which the absorption of one component is large compared to the sum of the absorptions of all other components (assuming all components are present in concentrations similar to those in an unknown mixture). In general, this condition is found at the position of maximum absorption of a band for one component.

There are some other factors to be considered in the selection of these wave-length positions. For example, in order to reduce the time necessary to make an analysis, it may be desirable to select bands near each other. It may be desirable to select positions which are outside of the water vapor region. If one thickness of the absorption

cell is to be used, the selection of the wave-length positions may depend upon the intensities of the bands; however, as far as possible, thinner absorption cells should be used when the bands are strong, and thicker ones should be used when the bands are weak. The selection of these positions is often complicated by the fact that several bands interfere with one another. This difficulty can be minimized by using the narrowest slits permitted by the sensitivity of the receiver and the available amplification of its output.

Determination of $\log I_0/I$ Values

For the measurement of $\log I_0/I$ values, the thermopile in the spectrometer is connected directly to a galvanometer with a sensitivity of one-tenth microvolt per millimeter at one meter.

TABLE I. Theoretical resolution of the spectrograph.

λ	Actual slit widths*	Effective slit widths	$\frac{\lambda}{d\lambda}$	$\frac{dn}{B-d\lambda}$
2 μ	0.04 mm	0.0097 μ	208	640
3	0.04	0.0117	257	440
4	0.05	0.0128	314	560
5	0.06	0.0118	424	640
6	0.07	0.0117	514	800
7	0.08	0.0117	600	960
8	0.09	0.0117	686	1080
9	0.11	0.0127	711	1160
10	0.13	0.0130	768	1240
11	0.16	0.0140	787	1560
12	0.20	0.0163	738	1720
13	0.24	0.0183	711	1800
14	0.35	0.0238	589	2080
15	0.55	0.0339	443	2240

* These slit widths produce half scale deflection.

TABLE II. Resolution for operating conditions.

Spectral position $\lambda(\mu)$	Actual slit width (mm)	Effective slit width (μ) ($\delta\lambda$)	Actual resolution $\lambda/\delta\lambda$	Prism resolution $B\delta n/\delta\lambda$
8.23	0.55	0.071	116	1100
8.55	0.60	0.075	114	1120
8.88	0.70	0.092	96	1152
10.50	0.90	0.085	123	1400
11.28	1.10	0.096	117	1600
9.13	0.7	0.080	114	1168
10.31	0.9	0.088	117	1336
11.89	1.2	0.100	119	1704
12.30	1.3	0.116	107	1752

The scale-to-galvanometer distance for the present work is five meters. With this system it is necessary to use wider slits and therefore to accept lower resolution than that realized with the amplifier, but the accuracy of measurements is much better. The actual resolutions for the slit widths used are shown in Table II.

When the procedure for a given mixture is developed, the slit widths for each spectral position are set to give approximately 400-mm deflection when a non-absorbing material (background cell) is in the light path. Of course, after the necessary calibrations have been made for the analysis of a particular mixture, the same slit widths must always be used for this analysis.

After the proper slit widths have been selected, the spectral positions are re-examined under the conditions which would be used in making the analyses, that is with wide slits and very slow rotation of the prism. These records for the hexanes are shown in Fig. 5. The upper line on each of the records shows the deflection of the galvanometer for the background cell. The bottom lines in the figures show the zero positions, and displacements along these lines in the vertical direction represent zero shift. The absorption curves lying between the background curve and the zero lines are as designated. Correct $\log I_0/I$ values can be obtained only if (1) the deflection of the galvanometer is a known function of the energy striking the thermopile and (2) all energy reaching the thermopile lies in the selected wave-length interval. In the present case, all tests have shown the galvanometer system to be linear. However, the energy reaching the thermopile actually includes some energy in addition to that from the wave-

length interval, $\Delta\lambda$. This wave-length interval is determined by the slit width and the region of the spectrum for which the spectrometer is set. The radiation in this wave-length interval which passes through the absorption cell and strikes the thermopile is the radiation which should be measured. Any other energy which causes a change in the galvanometer deflection (or other measuring instrument) when the shutter is opened is called false energy and its effect must be taken into account.

Most of this false energy is probably due to scattering from various parts of the spectrometer of radiation in the region 1 to 4μ which is the region of maximum intensity incident on the



FIG. 5. Portions of the spectra of the five hexanes.

Curve no.	Component	Position of band
11	n hexane	11.28 μ
12	3 methylpentane	10.50
13	2,3 dimethylbutane	8.88
14	2 methylpentane	8.55
15	2,2 dimethylbutane	8.23

entrance slit. This energy interferes mostly at long wave-lengths (6 to 15μ) where the true energy is small and the slits are wide. When working above 7μ most of the false energy can be cancelled out by using a mica shutter which transmits most of the radiation up to 5μ but none above 7μ . In this case the radiation below 5μ always passes into the spectrometer and is therefore included in the "zero" reading. Only the radiation of wave-lengths greater than 7μ is interrupted by the mica shutter. However, this method does not cancel out all the false energy; therefore, more refinement is necessary. A method of making partial correction for the remaining false energy has been established by using a piece of Cellophane which has two regions of 100 percent absorption, one at 9.15μ and one at 11.90μ , and obtaining the difference in the galvanometer deflections when the cells are in the light path with, first the Cellophane, and second the mica shutter. The amount of false energy when the background cell is in the light path is different from the amount when the liquid cell is in, but is practically the same for all liquids used in the liquid cell. It is proportional to I_0 at any given wave-length and therefore can be expressed as a percent of I_0 . The false energy values at the two bands in the spectrum of the Cellophane film have been plotted on a graph showing wave-length *versus* false energy and a straight line has been drawn between the points. (See Fig. 6.) Curve (1) is for the background cell and curve (2) is for the liquid cell with any of the nine pure components in it. At every spectral position and given slit width, I_0 is made the same for all observations. Therefore, the false energy corrections can be

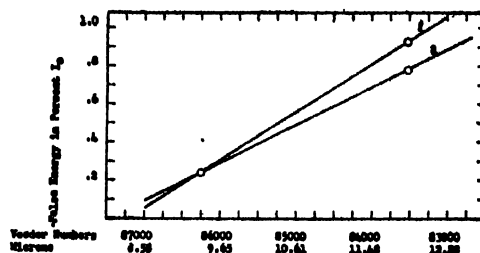


FIG. 6. False energy as a function of wave-length.

converted to mm deflection and subtracted from the I and I_0 values obtained, as shown in Table III.

With the equipment used in this work there is some zero drift and the intensity of the light source is not perfectly controlled. For this reason the procedure in making a measurement at a given spectral position consists of making three readings as follows: (a) with the absorption cell alone in the light path, (b) with the cell plus a mica shutter in the light path, and (c) with the absorption cell again alone in the light path.

The averaging of (a) and (c) in general cancels out the zero drift and the subtraction of this average from (b) gives (k). This number is proportional to the difference in the energy which falls on the thermopile with the two shutter positions. In other words, (k) is proportional to the energy which is interrupted by a mica shutter. A similar value (L) is obtained with the background cell. Then because the source may have changed its intensity during this interval of time, a value similar to (k), called (m), is again obtained. (k) and (m) are then averaged to give (n). This should, in general, cancel out any errors due to slow changes in the source. (n), therefore, is I plus the false energy F and

TABLE III. False energy corrections for each spectral position and slit width.

Principal absorber	$\lambda(\mu)$	cm^{-1}	Slit widths (mm.)	I_0 (mm)	I_0 (mm)	False energy correction I (mm)
2,3 dimethylbutene-1	9.13	1096	0.70	320.0	0.5	0.5
2,3 dimethylbutene-2	10.31	970	0.90	345.0	2.0	1.5
2,2,3 trimethylbutene	11.89	841	1.30	390.0	3.5	3.0
Methylpentenes	12.30	813	1.40	375.0	4.0	3.5
2,2 dimethylbutane	8.23	1215	0.55	450.0	0.5	0.5
2 methylpentane	8.55	1170	0.60	420.0	0.5	0.5
2,3 dimethylbutane	8.88	1126	0.70	495.0	0.5	0.5
2 methylpentane	10.50	952	0.90	430.0	2.0	2.0
n hexane	11.28	886	1.10	430.0	3.0	2.5

TABLE IV. Standard deviations of $\log I_0/I$ values for two samples.

Sample 223		Sample 756	
0.549	0.549	0.153	0.154
.550	.554	.153	.154
.548	.545	.153	.150
.548	.547	.152	.151
.548	.548	.150	.150
.549	.548	.152	.149
.547	.544	.149	.153
.551	.548	.155	.148
.550	.546	.153	.147
.552	.548	.145	.150
.551	.552	.149	.151
.547	.550	.150	.150
.549	.550	.151	.152
.546	.551	.151	.149
.548	.548	.149	.151
Average	.548		.151
Max. dev. from the mean			.006
Std. dev.	.0022		.0023

(L) is I_0 plus the false energy F_0 . These observations are written in a data sheet as follows:

Absorption cell			Background cell		
Shutter open	Shutter closed	Shutter open	Shutter open	Shutter closed	Shutter open
(a)	(b)	(c)	(d)	(e)	(f)
(g)	(h)	(j)			

The values of $I_0 + F_0$ and $I + F$ are obtained from the above readings in the following manner:

$$(b) - \left[\frac{(a) + (c)}{2} \right] = (k), \quad (e) - \left[\frac{(d) + (f)}{2} \right] = (L),$$

$$(h) - \left[\frac{(g) + (j)}{2} \right] = (m), \quad \frac{(k) + (m)}{2} = (n),$$

$$(L) = I_0 + F_0, \quad (n) = I + F.$$

F and F_0 are the false energy values which are subtracted to determine I_0 and I (see Table III).

It is essential in this method that changes in optical densities due to unavoidable changes in the two cells be known. It is, therefore, desirable to have a standard absorbing material which can be checked daily. This check is made with a "pure" sample of 2,2,3 trimethylbutene at 11.89 μ at the beginning and end of each day. When the $\log I_0/I$ deviates from the standard value (0.600) by more than 0.003, all values of $\log I_0/I$ determined on that day are corrected by adding or subtracting the indicated amount.

Reproducibility in Determining $\log I_0/I$ Values

Thirty determinations of $\log I_0/I$ values at one spectral position for two samples have been made over a period of three months. These values are listed in Table IV.

The standard deviations for the two sets of data are essentially the same even though the amount of absorption differs. For this particular case a change of 0.002 in $\log I_0/I$ represents a change of 0.3 percent in the concentration of the component sought. For a cell twice as thick it would represent 0.15 percent, for one-half as thick, 0.6 percent. Likewise the deviation in terms of percent concentration would be more or less, depending on whether the principal absorber absorbs more or less at this wave-length.

Analytical Curves

In making infra-red analyses it is necessary to convert values of $\log I_0/I$ to concentrations. The conversion may be made by means of analytical curves each of which shows $\log I_0/I$ as a function of concentration of a particular compound. In the present work with liquids, the data for such a curve are obtained by measuring $\log I_0/I$ with binary mixtures in the absorption cell.

The amount of absorbing material in the path is varied by using a cell of constant thickness and diluting the absorber with one of the compounds which has a high transmission at the wave-length in question. Three series of binary mixtures were made for the four-component mixture and four series were made for the hexane mixture. These solutions have concentrations as shown in Table V. It will be seen that for the four-component problem, one of the components of all the mixtures is 2,3 dimethylbutene-2. This was selected because it absorbs strongly at only one of the spectral positions. At all the other positions it absorbs very little. In the case of the hexane series 2,2 dimethylbutane was selected as the common component.

The preliminary analytical curves are obtained by plotting, for each spectral position, a curve showing percent concentration of the material diluted, which is 100 percent in the "a" solution (Table V), *versus* the $\log I_0/I$ for each of the binary mixtures (a through h) in the series. For the four-component problem, this means three

TABLE V. Compositions of binary mixtures used for analytical curves.

Mixture series no.	Components	Amounts by percent volume							
		a	b	c	d	e	f	g	h
Four component mixtures									
1	Methylpentenes	100	64	32	16	8	4	2	0
	2,3 dimethylbutene-2	0	36	68	84	92	96	98	100
2	2,3 dimethylbutene-1	100	64	32	16	8	4	2	0
	2,3 dimethylbutene-2	0	36	68	84	92	96	98	100
3	2,2,3 trimethylbutene	100	64	32	16	8	4	2	0
	2,3 dimethylbutene-2	0	36	68	84	92	96	98	100
Hexane mixtures									
11	Normal hexane	100	80	60	30	15	7	4	0
	2,2 dimethylbutane	0	20	40	70	85	93	96	100
12	3 methylpentane	100	80	60	40	20	10	5	0
	2,2 dimethylbutane	0	20	40	60	80	90	95	100
13	2,3 dimethylbutane	100	80	60	40	20	10	5	0
	2,2 dimethylbutane	0	20	40	60	80	90	95	100
14	2 methylpentane	100	80	60	40	20	10	5	0
	2,2 dimethylbutane	0	20	40	60	80	90	95	100

curves at each of the four spectral positions, and for the hexanes, four curves at each of the five spectral positions.

The final analytical curves are obtained from these preliminary curves by a procedure which will now be described. First, the final analytical curve of the component which has the least absorption at the spectral position being considered is drawn as a straight line from the (0.0) point to the point at 100 percent concentration. This is illustrated by Fig. 7 which is a graph of the preliminary and final analytical curves at 9.31μ for the four-component mixture. At this spectral position 2,2,3 trimethylbutene (C, Fig. 7) is the least absorbing pure compound. The contribution of this component to the $\log I_0/I$ value of the binary mixtures of this compound with 2,3 dimethylbutene-2 is determined from this final curve for 2,2,3 trimethylbutene. This contribution (lower C curve, Fig. 7) is subtracted from the preliminary curve of the binary mixture of these two components (upper C curve, Fig. 7) at a convenient number of points. These differences are the contribution to the $\log I_0/I$ values of 2,3 dimethylbutene-2 in each binary mixture and, when plotted against the concentration of 2,3 dimethylbutene-2, represent the final analytical curve of this component. As 2,3 dimethylbutene-2 is the common component in binary mixtures of each of the other compounds, the final analytical curves for the remaining two compounds can be obtained by a

series of subtractions similar to that described above. (Subtract difference between upper and lower C curves from upper A and D curves to obtain lower A and D curves). In this way an analytical curve for each of the components in the four-component mixture has been obtained

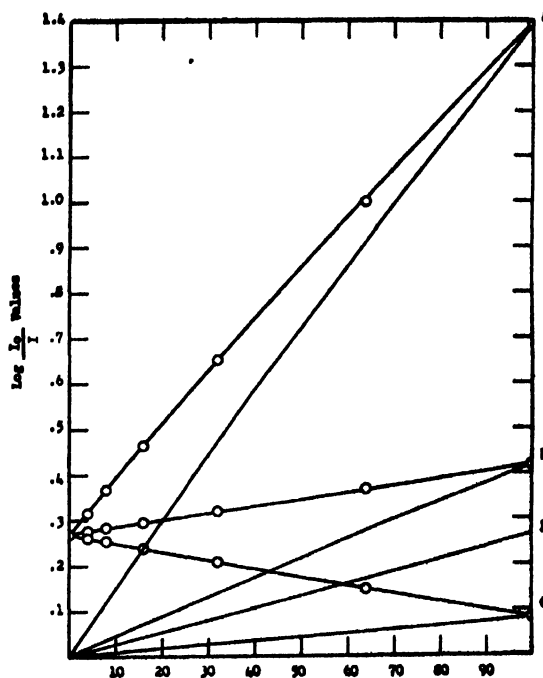


FIG. 7. Analytical curves at 9.31μ . Curves with circles are preliminary curves.

A. 2,3 dimethylbutene-1
B. 2,3 dimethylbutene-2
C. 2,2,3 trimethylbutene
D. methylpentene

at each of the four spectral positions (16 curves). For the hexane mixtures there are five curves at each of five spectral positions or a total of 25 curves.

After the analytical curves are obtained, it is possible to analyze mixtures which contain any or all of the components in any possible concentration. To do this, it is necessary only to obtain the $\log I_0/I$ values at all the spectral positions (n , if an n component mixture) and to convert them to concentrations by use of one of the two methods described below.

Two Methods of Converting Optical Densities to Concentrations

One method of converting optical densities to concentrations is a method of graphical approximation. In this case the analyses are arrived at by successive approximations based upon the final analytical curves. (See Fig. 7.) This method presupposes that the optical density of the mixture is equal to the sum of the optical densities of the component parts of that mixture as expressed in Eq. (3).

In converting optical densities to concentrations by this method, the $\log I_0/I$ value at one spectral position is applied to the analytical curve of the material which has its major absorption at that wave-length, and the first approximation of percent concentration for that component is read from the curve. The $\log I_0/I$ value at a second spectral position is corrected by subtracting the contribution of the first component at this wave-length. The corrected $\log I_0/I$ value is applied to the analytical curve of the material which has its major absorption at this position to obtain the first approximation for this component. The $\log I_0/I$ value at a third spectral position is corrected for the contribution of the first and second components and the corrected $\log I_0/I$ value is applied to the appropriate curve to obtain the first approximation for this third component. This procedure is continued until the first approximation is obtained for all components of the mixture. Successive approximations are obtained by repeating this process for each component, always using the latest approximate concentration of the other components. This is continued until two con-

secutive sets of approximations are the same. The last set of values is the analysis of the mixture.

The analytical curves can be used graphically, as just described, or they can be used to obtain mathematical expressions to make the same approximations. In this paper the latter will be called a method of applied corrections. The analytical curves, in general, are not straight lines (see Fig. 7). However, a straight line drawn between the points on the curve corresponding to concentration = 0 and concentration = 50 percent does not deviate far from the analytical curve at any point. Each of these straight lines is represented by an equation such as (1). The slopes of the straight lines are the approximate absorption coefficients, so that the optical density ($\log I_0/I$) of a mixture at a given spectral point can be written from Eq. (2) as:

$$(\log I_0/I)_{\lambda_j} = b_1c_1 + b_2c_2 + b_3c_3 + \cdots + b_nc_n, \quad (4)$$

where $b_1, b_2, \cdots, b_n$ are the slopes of the straight lines referred to above for components 1, 2, $\cdots, n$ at wave-length λ_j , and $c_1, c_2, \cdots, c_n$ are the concentrations. There is one of these equations for each of the 1, 2, 3, $\cdots, n$ wave-lengths.

For the four-component mixtures the equations are:

$$\begin{aligned} \left(\log \frac{I_0}{I}\right)_{\lambda_1} &= 0.01442c_1 + 0.00270c_2 \\ &\quad + 0.00840c_3 + 0.00499c_4, \\ \left(\log \frac{I_0}{I}\right)_{\lambda_2} &= 0.00193c_1 + 0.00391c_2 \\ &\quad + 0.00136c_3 + 0.00601c_4, \\ \left(\log \frac{I_0}{I}\right)_{\lambda_3} &= 0.00155c_1 + 0.00082c_2 \\ &\quad + 0.00622c_3 + 0.00757c_4, \\ \left(\log \frac{I_0}{I}\right)_{\lambda_4} &= 0.00080c_1 + 0.00125c_2 \\ &\quad + 0.00092c_3 + 0.02792c_4, \end{aligned} \quad (5)$$

where the λ 's and c 's are as follows:

$$\begin{aligned} \lambda_1 &= 9.12\mu & c_1 & \text{is the concentration of} \\ & & & \text{2,3 dimethylbutene-1,} \\ \lambda_2 &= 10.31\mu & c_2 & \text{is the concentration of} \\ & & & \text{2,3 dimethylbutene-2,} \end{aligned}$$

$\lambda_3 = 11.89\mu$ c_3 is the concentration of
2,2,3 trimethylbutene,
 $\lambda_4 = 12.30\mu$ c_4 is the concentration of
methylpentenes.

With considerable effort the above four equations can be solved for the concentrations, giving Eqs. (6).

$$\begin{aligned} c_1 &= 76.28 \left(\log \frac{I_0}{I_1} \right) - 52.72 \left(\log \frac{I_0}{I_2} \right) \\ &\quad + 1.335 \left(\log \frac{I_0}{I_3} \right) - 0.7350 \left(\log \frac{I_0}{I_4} \right), \\ c_2 &= -32.10 \left(\log \frac{I_0}{I_1} \right) + 304.0 \left(\log \frac{I_0}{I_2} \right) \\ &\quad - 55.45 \left(\log \frac{I_0}{I_3} \right) - 45.47 \left(\log \frac{I_0}{I_4} \right), \\ c_3 &= -14.48 \left(\log \frac{I_0}{I_1} \right) - 12.52 \left(\log \frac{I_0}{I_2} \right) \\ &\quad + 171.7 \left(\log \frac{I_0}{I_3} \right) - 41.67 \left(\log \frac{I_0}{I_4} \right), \\ c_4 &= -0.2710 \left(\log \frac{I_0}{I_1} \right) - 11.69 \left(\log \frac{I_0}{I_2} \right) \\ &\quad - 3.214 \left(\log \frac{I_0}{I_3} \right) + 39.19 \left(\log \frac{I_0}{I_4} \right). \end{aligned} \quad (6)$$

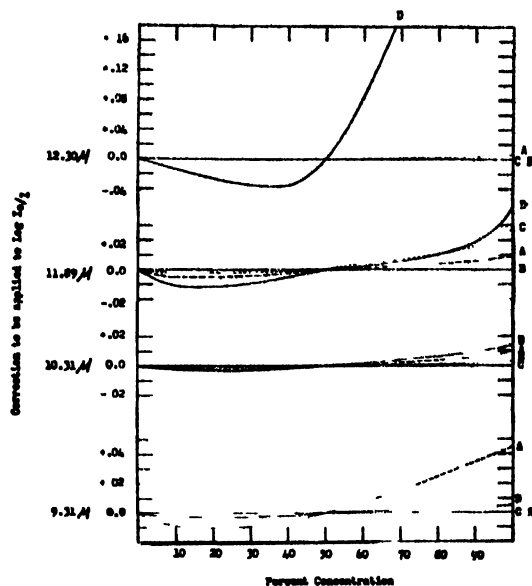


FIG. 8. Correction curves for four-component mixture used with method of applied corrections. (See Fig. 7 for explanation of A, B, C, and D.)

Equations (6) are used in connection with a series of correction curves which are obtained in the following manner: the differences between the $\log(I_0/I)$ values as obtained from the analytical curves and the straight lines mentioned above are plotted against the concentrations of each component at each wave-length. For the four-component mixture, these are shown in Fig. 8.

TABLE VI. Results of analyses of synthetic mixtures of four components.

Sample no	2,3 dimethylbutene -1		2,3 dimethylbutene -2		2,2,3 trimethylbutene		Methylpentenes	
	Actual	Spec.	Actual	Spec.	Actual	Spec.	Actual	Spec.
U-1	25.0%	25.2%	25.0%	23.2%	25.0%	25.7%	25.0%	24.9%
U-2	30.0	30.4	20.0	18.5	30.0	31.4	20.0	19.9
U-3	20.0	20.1	30.0	29.3	35.0	36.9	15.0	15.1
U-5	33.3	33.2	33.3	33.9	33.3	33.6	0.0	0.0
U-6	33.3	34.0	33.3	33.5	0.0	0.0	33.3	33.3
U-7	0.0	0.0	33.3	34.0	33.3	34.3	33.3	33.6
U-8	33.3	33.5	0.0	0.0	33.3	34.3	33.3	33.7
O-1	25.0	24.6	25.0	23.9	25.0	24.4	25.0	25.2
O-2	12.5	11.5	62.5	63.5	12.5	11.2	12.5	12.3
O-3	42.0	42.8	18.0	17.3	20.0	19.0	20.0	19.4
O-4	36.0	35.6	18.0	16.9	26.0	25.1	20.0	20.5
O-5	33.3	33.6	33.3	31.9	33.3	32.6	0.0	0.0
O-6	33.3	33.6	0.0	0.0	33.3	33.5	33.3	33.0
O-8	33.3	33.1	33.3	33.0	0.0	0.3	33.3	33.0
O-9	50.0	50.7	0.0	0.0	50.0	50.1	0.0	0.0
O-10	0.0	0.0	0.0	1.1	50.0	49.7	50.0	49.9
O-11	2.5	1.9	3.0	3.0	90.0	88.8	4.5	4.9
Av. dev.	0.36%		0.72%		0.77%		0.25%	
Max. dev.	1.0		1.8		1.9		0.6	

TABLE VII. Results of analyses of synthetic mixtures of five components.

Sample no.	2,2 dimethylbutane		2,3 dimethylbutane		2 methylpentane		3 methylpentane		n hexane	
	Actual	Spec.	Actual	Spec.	Actual	Spec.	Actual	Spec.	Actual	Spec.
H-1	20.0%	20.4%	20.0%	19.8%	20.0%	19.2%	20.0%	19.6%	20.0%	19.8%
H-2	10.0	10.4	10.0	10.1	10.0	9.7	10.0	9.6	60.0	59.2
H-3	5.0	5.4	5.0	5.3	5.0	5.0	55.0	54.0	30.0	28.4
H-4	2.5	2.8	52.5	51.8	2.5	2.2	27.5	27.1	15.0	15.1
H-5	1.2	1.6	26.3	25.9	51.2	50.3	13.8	13.1	7.5	7.8
H-6	50.6	50.7	13.1	13.3	25.6	24.7	6.9	6.8	3.8	4.4
Av. dev.	0.33%		0.32%		0.54%		0.50%		0.60%	
Max. dev.	0.4		0.7		0.9		1.0		1.6	

To make an analysis the log (I_0/I) values of the mixture are determined at each of the selected spectral positions and are substituted in Eqs. (6). The first approximate concentrations are obtained from these equations. The correction to be applied to the experimentally determined log (I_0/I) at each wave-length position is determined by subtracting the sum of the values obtained from the correction. These new log (I_0/I) values are then used in Eq. (6) and second approximate concentrations are determined. This process is carried on until the con-

centrations are the same for two consecutive approximations. The last set of values is the analysis. In the present work, three approximations have been sufficient.

DISCUSSION OF DATA

In order to check the methods, synthetic mixtures have been made and analyzed. The results are shown in Table VI for the four-component mixture and in Table VII for the hexanes. It will be seen that in no case does the analysis obtained deviate from the actual composition

TABLE VIII. Comparison of results by two methods.

Sample no.	2,3 dimethylbutene -1		2,3 dimethylbutene -2		2,2,3 trimethylbutene		Methylpentenes	
	Graphical approximation	Applied correction	Graphical approximation	Applied correction	Graphical approximation	Applied correction	Graphical approximation	Applied correction
O-3	20.1%	19.8%	29.3%	29.1%	36.9%	37.0%	15.1%	15.1%
O-4	35.6	35.5	16.9	16.6	25.1	24.7	20.5	20.4
O-5	33.7	33.6	32.2	31.9	32.5	32.6	0.0	0.0
O-11	2.0	1.9	2.9	3.0	89.2	88.8	5.0	4.9

TABLE IX. Data showing the effects of extraneous elements.

Sample		2,3 dimethylbutene -1	2,3 dimethylbutene -2	2,2,3 trimethylbutene	Methylpentene	2 methylbutene -2	Octene cut
U-5	As made	33.3%	33.3%	33.3%	0.0%	0.0%	0.0%
	Analysis	33.2	33.9	33.6	0.0	Assumed Absent	
U-7	As made	32.5	32.5	32.5	0.0	2.5	0.0
	Analysis	33.4	35.8	28.1	2.4	Assumed Absent	
U-8	As made	0.0	33.3	33.3	33.3	0.0	0.0
	Analysis	0.0	34.0	34.3	33.6	Assumed Absent	
U-8	As made	0.0	32.5	32.5	32.5	0.0	2.5
	Analysis	0.0	33.5	33.5	32.6	Assumed Absent	
U-8	As made	33.3	0.0	33.3	33.3	0.0	0.0
	Analysis	33.5	0.0	34.3	33.7	Assumed Absent	
U-8	As made	31.5	0.0	31.5	31.5	2.5	2.5
	Analysis	35.9	0.0	34.4	34.0	Assumed Absent	

by more than 2.0 percent of the total mixture. The average error for all components in all mixtures shown in the two tables is 0.5 percent.

All the synthetic mixtures have been analyzed by the graphical approximation method, and some have been analyzed by both methods. Results obtained by the two methods for four samples are compared in Table VIII.

The complete analysis of a four-component mixture can be made in 45 minutes with the method of graphical approximations, and in 30 minutes with the method of applied corrections.

It should be pointed out that the presence in the unknown mixture of compounds for which analyses are not made will cause large or small errors depending upon whether the extraneous materials absorb strongly or weakly at any of the spectral positions. It has been found in the case of the four-component mixture that many of the unknowns submitted for analysis have contained two extraneous compounds. Synthetic mixtures including 2.5 percent of each of these components were prepared and analyzed, assuming the extraneous components absent. The results of this study are shown in Table IX. It

will be noticed that these concentrations of the extraneous components cause errors up to 4.0 percent in the analyses. If the approximate amount and effect of the extraneous components can be determined, and they frequently can, errors such as those shown in the example can be eliminated or at least minimized by making appropriate corrections.

SUMMARY

A method has been described for analyzing multicomponent mixtures of hydrocarbons in the liquid phase by means of their infra-red absorption spectra. A constant thickness absorption cell is used, and analytical curves are established on the basis of synthetic binary mixtures of the different components in the multicomponent mixtures to be analyzed. Two procedures for using the analytical curves to convert optical densities to concentrations by means of successive approximations have been described. Four- and five-component mixtures are completely analyzed in less than 45 minutes per sample and the average error is 0.5 percent of the total sample.

Wide Range Static Strength Testing Apparatus for Glass Rods*

T. C. BAKER† AND F. W. PRESTON
Preston Laboratories, Butler, Pennsylvania
(Received November 1, 1945)

By adapting an electro-dynamic loudspeaker, an apparatus was assembled to test the strength of glass rods in flexure. This apparatus will apply the stress in 2 to 3 milliseconds and permit static loading for times as short as 0.01 second. Using an additional long time loading apparatus, which simultaneously tested eighteen rods, it was possible to run static fatigue tests on glass rods from 1000 seconds to 86,400 seconds. The method of testing and errors due to the apparatus are discussed.

INTRODUCTION

WHENEVER an attempt has been made to measure the strength of glass, it has been found that the rate of application of the stress greatly affects the results. If the load applied to the specimens is increased continuously, it is found that the specimens that are loaded slowly break at much lower stresses than those which are loaded rapidly. If the load is applied by increments, with definite time intervals between each addition, the specimens which are loaded with long waits between steps break at much lower stresses than those loaded with short waits between the steps. Moreover, a specimen under constant load may survive several hours before it breaks. This apparent weakening of glass with increased time of loading has herein been denominated "fatigue." Since engineers commonly use the term "fatigue" to mean failure under oscillatory or repeated loads, we may denote the fatigue dealt with here as "static" fatigue, and the other kind as "dynamic" fatigue.

Naturally it is important to know, for both theoretical and commercial reasons, what the real nature of fatigue is. The theoretical questions are: what is the strength of glass for very long loading periods, and does its instantaneous strength approach the value suggested by molecular theory? The commercial questions are: what safety factors should be allowed for bottles containing pressure, and what impact stresses will a glass object stand? Although there had been considerable speculation on these questions, the data were so limited as to make no generalization really sound. The object of this work was the

extension of the data by several orders of magnitude to both shorter and longer times.

Probably the most relevant data would show how long a specimen survives at a given stress. Unfortunately, to answer this question directly with experimental methods would be too tedious, but the inverse question, whether or not the specimen survives a given length of time at a predetermined stress, lends itself readily to experimental technique. After sufficient data had been collected to answer this question, plotted curves would provide the answer to the original question. Accordingly, the method finally decided upon was to load the test specimen for a definite interval of time at a predetermined stress. If the specimen survived the test, it would be loaded at an appreciably higher stress and the test repeated.

Round glass rods were chosen as the only satisfactory test specimens because of their availability in several compositions, and because they have no cut or ground surfaces to confuse the issue.

When a constant load is to endure for only 0.01 second, the problem of applying the load and removing it is far more difficult than when the load is to be held constant for hours, because for the brief application the load must be built up with great speed and without producing oscillation. Therefore, two types of apparatus were designed: a short time loading apparatus that would apply the load for periods ranging from 0.01 second to 100 seconds, and a long time loading apparatus that would apply the load for periods ranging from 1000 seconds to 24 hours (86,400 seconds).

DESCRIPTION OF APPARATUS

The Short Time Loading Apparatus

The short time loading apparatus is essentially a dynamic motor consisting of a modified radio

* This apparatus was developed at the Preston Laboratories in 1938 and 1939 for a project sponsored by the Hartford-Empire Company.

† Present address: Libbey-Owens-Ford Glass Company, Plaskon Division, Toledo, Ohio.

loudspeaker. Instead of the usual voice coil, an armature or moving coil assembly was developed which would withstand the loads that had to be applied.

A schematic diagram of the apparatus is shown in Fig. 1. A glass rod (1) is supported by two knife edges (2). A wire bail (3) is looped over the middle of the glass rod and is connected to two lugs (4) on the armature (5). An oversize field coil (6) which is actuated by a 12-volt storage battery (7) produces the magneto-motive force of a central field magnet (8). The magnetic circuit is carried from the central magnet through a shell (9) to the air gap (10), in which the armature (5) is freely suspended from the rod. When the key (11) is closed, the armature is actuated by the current flowing from the battery (12) of about 120 volts. The current is regulated by a resistance (13) and is measured by an

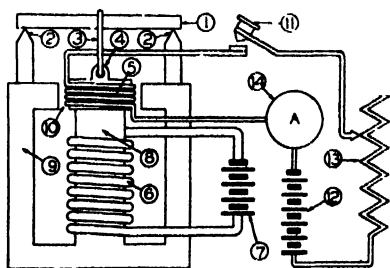


FIG. 1. Schematic diagram of the quick-loading apparatus.

ammeter (14). Forces up to 30 lb. can be produced momentarily by the armature.

Details of the Dynamic Motor

An eighteen-inch dynamic loudspeaker was used as the basis of the dynamic motor. The size of the field coil was practically doubled to provide a much stronger magneto-motive force. The magnetic air gap was operated at approximately 10,000 gauss for the smaller armatures, and as low as 7,000 gauss for the larger ones. In Fig. 2 the rod is shown in position to be tested. Several armatures were designed for different purposes. For the 0.01-, 0.1-, and 1-sec. loading durations the requirements were light weight and large damping coefficient to permit quick action without overtravel. The time interval was so short that the motor coil could carry large currents

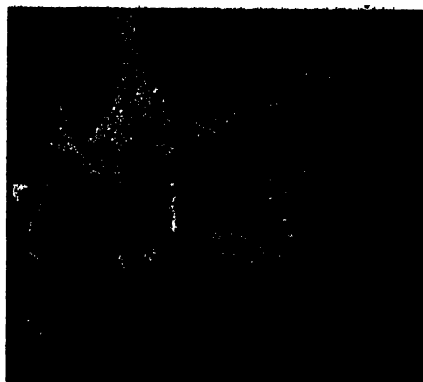


FIG. 2. A rod in position to be tested.

without overheating. However, for the 10-second and 100-second periods the chief requirement was that the coil should not overheat, necessitating a larger motor coil and greater weight, yet the time of application of the load could be much longer so that smaller damping coefficients and greater weight could be tolerated. The load could be satisfactorily applied by reducing the circuit resistance by hand, thus bringing the current to its correct value in approximately $\frac{1}{4}$ second.

Figure 3 shows a cross-sectional diagram of one side of the armature. An aluminum form serves as a frame and a spool. The wire bail to the rod is fastened at the hole (1) in the lug (2). A stiffening ring (3) prevents the form from collapsing due to the pull of the wire bail. A motor winding (4), which carries the current from the 120-volt battery, consists of a double layer of No. 28 wire. The damping winding (5) is made of several turns of fine, uninsulated, aluminum wire. However, the damping winding is insulated from the aluminum spool and from the motor winding by



FIG. 3. Cross section of voice coil.

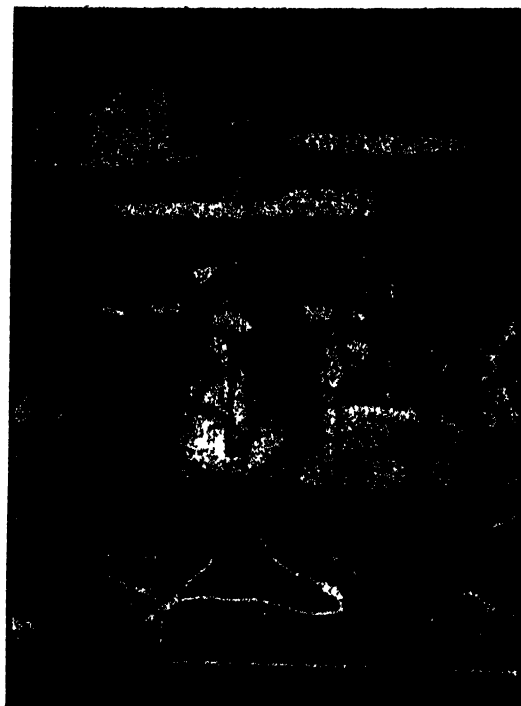


FIG. 4. Trip scale for calibrating the dynamic motor

a layer of insulating enamel. It was possible to thicken the spool and omit the damping winding in the design of some of the armatures.

The operation of the dynamic motor depends upon the fact that a force is exerted on a wire carrying current in a magnetic field. This force is given by $F = BIl$, where F is the force, B is the flux density in gauss, l is the length of conductor in the field, and I is the current flowing through the wire, in c.g.s. electromagnetic units. Thus the force is directly proportional to the current in the motor winding. As long as there are approximately the same number of turns of wire in the magnetic field, the force will be substantially constant for small displacements of the armature. In practice, the armature can be displaced over a range of $\frac{1}{16}$ inch while the force remains constant to $\pm \frac{1}{2}$ percent. The quick-loading armature designed for 0.01-second loading produces a force of 3 pounds per ampere.

When the motor winding moves across the magnetic field, a back e.m.f. is developed. This back e.m.f. is given by $E = Blv$, where E is the electromotive force in c.g.s. electromagnetic units, l is the length of conductor in the magnetic field, B is the flux density, and v is the velocity.

Thus, when the coil moves, it is possible to determine its velocity by means of an oscillograph and so tell whether it is behaving as it should.

The damping of the armature is due to the counter currents induced in the ring and damping winding by the motion of these conductors in the magnetic field. By definition, the coefficient of damping is $k = F/v$ where k is the damping constant, v is the velocity, and F is the force produced in opposition to the velocity. By combining this with the formulas $F = BIl$, and $E = Blv$, it turns out that $k = B^2 l^2 I / E = B^2 l^2 / R$. Here R is the electrical resistance of the conductor in the circuit. In the present case the damping constant k depends on three factors: the damping due to the aluminum spool, that due to the motor winding, and that due to the damping winding. Each of these factors can be computed from the formula $k = B^2 l^2 / R$. The coefficient of damping in the spool is the most difficult to determine because the path of the current extends beyond the boundaries of the uniform magnetic field. The currents in the damping winding and in the motor winding, on the other hand, are confined within the part of the field where the geometry is relatively simple. The three separate coefficients have to be added together to obtain the total damping. When an armature is made, a good estimate of the actual value of the coefficient can be obtained by studying the voltages across the motor coil with the aid of an oscillograph, as will be explained later. Corrections can be made by adding to or subtracting from the damping winding enough wire to give the desired coefficient.

The contribution of the motor winding to the damping constant is rather small if the circuit resistance stays fairly large. With a field strength of 10,000 gauss, and with 50 ohms in the circuit, the contribution is 4 lb./sec., and with 20 ohms in the circuit the contribution is 10 lb./sec. The circuit resistance was usually at least 20 ohms. This variation of 6 lb./sec. will change the total damping coefficient by only 3 or 4 percent.

A preload is applied to the armature by passing a small current through the motor coil in order to take the slack out of the wire bail. This preload puts a maximum stress upon the specimen of around 1000 to 2000 lb./sq. in. This is sufficiently below the breaking stress at any of the durations, and under any of the conditions where it is used,

that it causes no appreciable fatigue. This matter will be discussed more fully later.

Several armatures were designed for quick loading, because none is effective for more than a small range of rod sizes. For example, to test the rods that vary in diameter from 0.210 inch to 0.230 inch, a coil assembly was designed which has a damping coefficient of 168 lb./sec. and which weighs 23 grams.

For the 10- and 100-second loads a six-layer coil was designed which provides adequate loads at current values low enough to prevent overheating. The additional weight and reduced damping of this coil offer no problem at the relatively slow loading rates appropriate for these long loading periods.

Calibrating the Dynamic Motor

In order to measure the force exerted by the armature, a trip scale is mounted with the center of one pan over the motor coil. Figure 4 shows the trip scale in position. The wire bail of the armature is attached below the empty pan, and the opposite pan is loaded with a ten-pound weight. The empty pan is held down by hand, keeping the armature in its working range, while the current is increased until the coil just supports the ten-pound weight. Such determinations can be made to about $\frac{1}{2}$ percent accuracy with an ammeter readable to about $\frac{1}{4}$ percent of full scale. A better ammeter would be of no value as there is sufficient damping and friction in the dynamic motor to make finer readings uncertain. With a satisfactory ammeter only one calibration point is necessary, since the relation between the current through the motor coil and the force exerted by it is as linear as can be measured, if the field magnet is properly saturated. Of course, the current through the field coil must be kept constant or the calibration will change.

The Timing Device

The timing device which was designed to time the 0.01-second, 0.1-second, and 1.0-second intervals appears in Fig. 5. It consists of a 100-lb. pendulum mounted on a heavy frame and has a period of 2 seconds. The pendulum is released by a magnet and is caught at the end of its swing by a latch. At the start of the swing a limit switch is closed to energize the field magnet; at the end

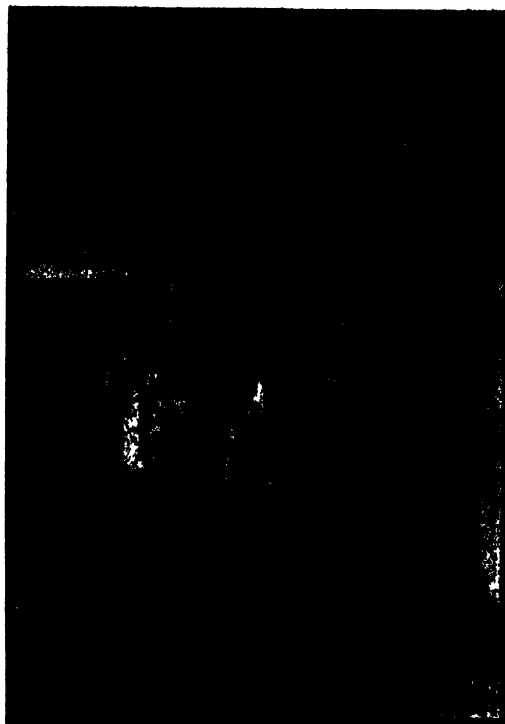


FIG. 5. The timing device.

of the swing another limit switch breaks the field coil circuit when the test cycle is complete.

The 0.01-second and 0.1-second intervals are controlled by means of a cam on the pendulum bob, which closes a pair of contacts in the middle of the swing. At the end of the interval another cam opens the contacts, and the resulting arc is blown out with a powerful jet of CO_2 . The arc is greatly diminished in about 0.0001 second, but it takes about 0.001 second to blow it out. Without the blast of CO_2 it is nearly impossible to put out the arc at all because the contacts open only $\frac{1}{4}$ inch and currents up to 6 amperes at 120 volts d.c. have to be interrupted. The timing cams are calibrated by observing the oscillograph pattern and making the indicated adjustments.

The 1.0-second interval is timed by having the pendulum close a switch in the middle of its forward swing, and open it on the return swing. Since the pendulum has a 2-second period, this interval is one second.

The 10- and 100-second intervals are timed by a stopwatch. A slidewire rheostat shown in Fig. 2 and indicated by (13) in Fig. 1 is operated by hand to apply and release the load.

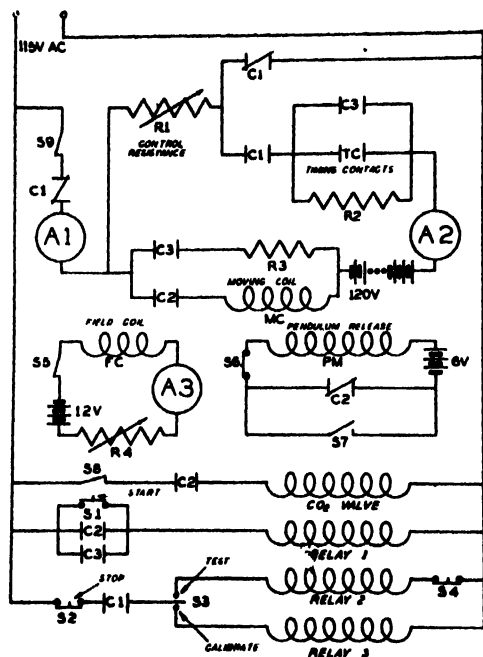


FIG. 6. The electrical control circuit.

Accessories to the Apparatus

Before many tests had been made, it became evident that a great deal of accessory apparatus was necessary. While the current in the loading circuit is being adjusted preparatory to a test, a dummy resistance is substituted for the motor coil in order to protect the latter from overheating. This resistance has to be switched in when the current is adjusted and out when a rod is tested. The control resistance becomes heated during the loading current adjustment, so that an a.c. current must be supplied to it between times to keep it warm and maintain its resistance constant. The CO_2 used to blow out the arcs of the timing contacts is turned on and off with an electric valve, because otherwise the flow would quickly dissipate the supply.

It was soon found that the manual switching arrangement was too complicated. Turning the wrong switch could easily cause an accident or spoil the test. For example, the pendulum bob had to be released when testing but not when calibrating. An interlocking circuit was finally evolved which was relatively foolproof and which reduced the switching procedure to a minimum. This circuit is shown in Fig. 6, and is described below.

It has been found that a cathode-ray oscillograph is indispensable. The timing cams are adjusted with the oscillograph, and the trace of every test is watched on the oscillograph screen to be sure that everything goes correctly. By watching the voltage developed across the motor winding as indicated by the oscillograph it is possible to tell whether the load goes on smoothly, whether the armature sticks, whether the rod breaks, or whether the contacts arc excessively.

Operation of the Apparatus

Because of the fact that the rods to be tested vary in diameter, the experimental procedure is somewhat complicated. First it is necessary to construct a chart giving the relationship between the maximum fiber stress of the rod and the current through the motor winding for every rod diameter. Each rod is carefully measured for maximum and minimum diameter and then is oriented in the apparatus so that the maximum fiber stress will come at one end of the minimum diameter.

Referring to Fig. 6 the sequence of operations is as follows: the knife switch (S5) is closed to energize the field magnet, and the manual switch (S3) is set on "calibrate"; the start button (S1) is pressed, and the current indicated by the ammeter (A2) is set to the proper value (as given by the stress vs. current table) by adjusting the control resistance (R1). Finally, to test the specimen, the switch (S3) is set on "test," and the start button (S1) is pushed again. At this



FIG. 7. Theoretical curve of load of 0.01-second duration for rods 0.230 and 0.210 inch in diameter.

point it may be helpful to trace on Fig. 6 the sequence of part of the resulting operations.

Closing (S1) energizes the coil of relay (1) which causes an open contact (C1) of this relay to close and energize the coil of relay (2). Other contacts (C1) open to turn off the a.c. supply maintaining the temperature of the control resistance (R1) and close to make the moving coil circuit ready for closure of the timing contacts (TC). The coil of relay (2) being energized by closure of contacts (C1), the contact points (C2) open in the pendulum release circuit, allowing the pendulum to fall. Thus the pendulum bob is released, and by closure of the open contacts (C1) of relay (1) the preload controlled by (R2) is turned on, the CO₂ valve is opened, and the timing contacts (TC) become operative. The pendulum bob swings down and operates the timing contacts (TC). This energizes the moving coil (MC) of the armature from the 120-volt battery, and loads the specimen. At the end of its swing, the pendulum bob operates a limit switch (S4) which shuts off the apparatus while the pendulum is caught by a mechanical latch. In the meantime the operator watches the oscillograph to see that everything has functioned properly. If the rod has not broken, the pendulum is reset, the current adjusted for a higher stress, and the whole procedure repeated.

Dynamics of Quick Loading

In order to get some idea of the accuracy of the measurements it is necessary to study the loading mechanism as a vibrating system. Since most of the mass is in the armature and most of the compliance is in the rod, the problem may be treated as a vibrating system with a single degree of freedom. The differential equation becomes:

$$m \frac{d^2 y}{dt^2} + k \frac{dy}{dt} + sy = F,$$

where y is the displacement of the center of the rod, m is the effective mass of the whole system, k is the damping constant, s is the stiffness of the rod at its center, and F is the suddenly applied force. The load F is applied at $t=0$, so the equation is subject to the boundary conditions:

$$y=0 \text{ and } dy/dt=0 \text{ when } t=0.$$

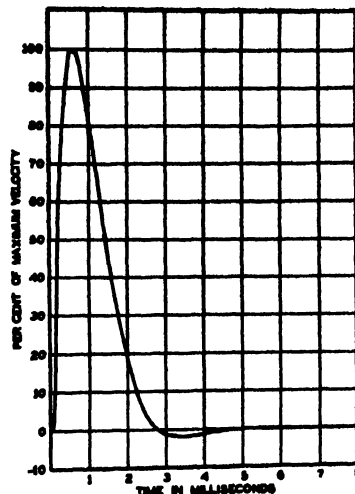


FIG. 8. Theoretical armature velocity curve during application of the load.

The solution of the equation becomes:

$$y = \frac{F}{s} - \frac{F}{s \cos \phi} e^{-kt/2m} \cos(\omega t - \phi),$$

which is plotted for two rod diameters in Fig. 7. Then

$$\frac{dy}{dt} = -\frac{F}{s \cos^2 \phi} e^{-kt/2m} \sin \omega t,$$

which is plotted in Fig. 8. In these equations:

$$\cos \phi = \frac{\left(\frac{s}{m} - \frac{k^2}{4m^2} \right)^{1/2}}{(s/m)^{1/2}} \quad \text{and} \quad \omega = \left(\frac{s}{m} - \frac{k^2}{4m^2} \right)^{1/2}.$$

As shown in Fig. 7, the displacement is minimum at $t=0$ and passes through a maximum at 2.8 milliseconds or, in general, at $t=\pi/\omega$.

Likewise, as seen in Fig. 8, the velocity passes through a maximum followed by a minimum, and these occur at $t=(\pi/2-\phi)/\omega$ for the maximum and $t=(3\pi/2-\phi)/\omega$ for the minimum.

These correspond to the maximum and minimum e.m.f.'s across the motor coil as indicated at D and G in Fig. 9, which is described later.

The ratio between the overtravel, i.e., the excess displacement over that for the steady load, and the steady load displacement is $e^{-\pi \tan \phi}$. The ratio of the minimum velocity to the maximum velocity is also $e^{-\pi \tan \phi}$. Therefore, because the voltage induced in the moving coil is proportional

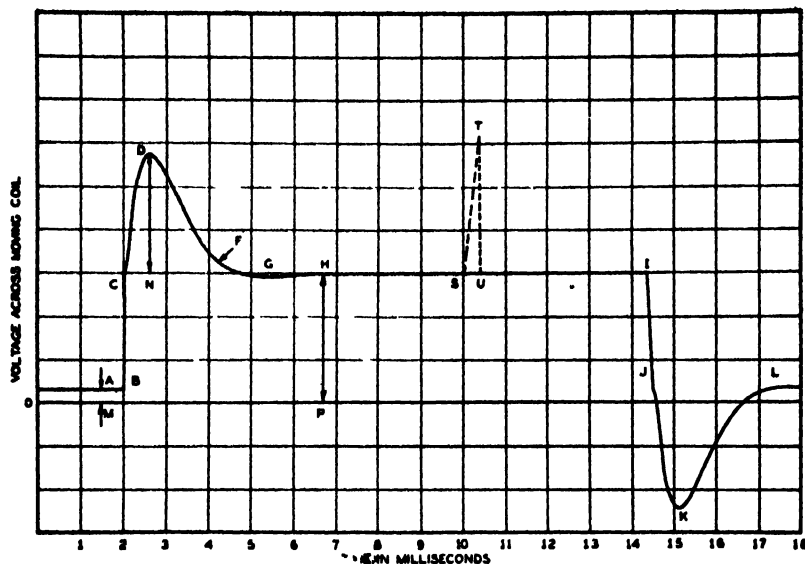


FIG. 9. Oscillograph trace produced during a 0.01-second test.

to its velocity, $e^{-\pi \tan \phi}$ must equal the ratio of the departures from the steady load voltage at D and G in Fig. 9. By measuring this ratio in the actual tests, the amount of overtravel and the damping coefficient can be readily computed. It turns out that for 2 percent overtravel, $k = 1.20(ms)^{1/2}$. For a range of rod sizes to be tested with an armature of given mass and damping coefficient, the largest diameter is determined by the condition that there should be no more than 2 percent overtravel. The smallest diameter is determined by not allowing the duration of the load to be decreased by more than 10 percent for the 0.01-second test.

Originally it was intended to load the rods in one millisecond, which would require the use of rods greater than $\frac{1}{4}$ inch in diameter, with a 4-inch span between supports. Unfortunately, the rather surprising increase of strength of glass in the shorter time ranges completely overtaxed the equipment with the larger rods. It would have required a larger field magnet, much higher d.c. voltages, and a general redesign of the apparatus. So we have had to use rods 0.220 ± 0.010 inch in diameter with a 5-inch span, and to be content with loading the rods in 2 to 3 milliseconds instead of the one millisecond originally planned. However, as one can see from Fig. 7, the time the rods are under stresses between 80 percent and 98 percent of the full loads is not too great.

A typical armature designed for measuring

rods of approximately 0.220-inch diameter has the following constants:

$$m = 0.051 \text{ lb.}, \quad k = 168 \text{ lb./sec.}$$

It turns out that this coil causes a momentary 2 percent overtravel on rods of 0.230-inch diameter, and it can be used with rods of 0.210-inch diameter without decreasing the time of full loading by more than 10 percent. Figure 7 shows a theoretical loading curve for rods 0.230 inch and 0.210 inch in diameter. The constants of the system with the 0.230-inch diameter rod are:

$$m = .060 \text{ lb.}, \quad k = 168 \text{ lb./sec.},$$

and

$$s = 194,000 \text{ lb./sec.}^2.$$

The mass constant is increased over the mass of the armature by the effective mass of the rod.

The constants for the rod 0.210 inch in diameter become:

$$m = .059 \text{ lb.}; \quad k = 168 \text{ lb./sec.},$$

and

$$s = 135,000 \text{ lb./sec.}^2.$$

In Fig. 8 is a theoretical curve of the velocity of the armature for a rod 0.230 inch in diameter. It turns out that this pattern is closely followed when the velocity is studied in the cathode-ray oscillograph.

Oscillograph Patterns

Since the velocity of the armature produces an

e.m.f. given by

$$E = Blv, \text{ and since } v = dy/dt, \quad \bullet$$

then

$$E = Bl \frac{F\omega}{s \cos^2 \phi} e^{-kt/2m} \sin \omega t,$$

and we should expect the voltage developed across the armature to have this e.m.f. as one component. The other component is, of course, the IR drop across the coil. Accordingly, when the oscillograph is connected to read the voltage across the armature, a pattern is given as shown in Fig. 9. The line $ABCDIJKL$ is a trace of the oscillograph in a 0.01-second test. AM is the IR drop due to the preload current; HP is the IR drop due to the full load current. The voltage DN is the maximum e.m.f. due to the velocity. The load comes on at about F , and the overtravel occurs in the region between F and G . The time the load is on is between F and I . If a rod breaks, a pattern results as shown by the dotted line STU .

It is thus quite easy to determine from the oscillograph pattern whether the load has been properly applied, how long the load was on, whether the rod broke, and whether or not any deleterious accident, such as the armature sticking, took place during the test. It is somewhat difficult to estimate the amount of overtravel by increasing the scale, because the IR drop is considerably larger than the e.m.f. due to velocity in the 2 percent overtravel region. It is found easier to use an auxiliary winding of a few turns of No. 40 wire around the outside of the armature when damping coefficients are being estimated by the overtravel.

Accuracy of Tests

The quick-loading device is able to apply a load in 2 to 3 milliseconds, maintain it for 10 milliseconds, and then remove it in $\frac{1}{2}$ millisecond. Even when a range of rod sizes is tested, the error in timing the load is less than ± 5 percent, and the overtravel is less than 2 percent. Since the overload is on for only about 10 percent of the time it is doubtful that a 2 percent overload would cause as much as 1 percent difference in the results, because of the rapid increase of strength of glass with shortening of duration of loading. An added factor in favor of accuracy is

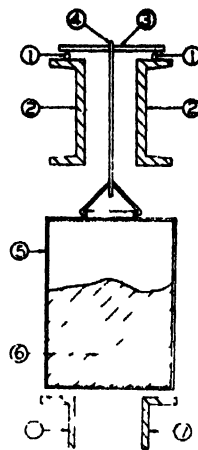


FIG. 10. Cross section of the long time loading device.

that, when a large number of rods are tested, several will not be overloaded, because they are on the low side of the diameter range. Therefore, for the 0.01-second test, the average result can be relied upon to at least 1 percent as far as the overloading is concerned.

The overload will not be serious in the 0.1-second and 1-second tests, because it will occur for only about 1 percent and 0.1 percent of the total time under load. The manner of application of the 10- and 100-second loads makes overloading in these tests impossible. Thus it is safe to say that 1 percent accuracy for the load and ± 5 percent accuracy for the timing of the load are available for all the short-time loading points.

The Long Time Loading Device

The problem in the long time loading device was to test several rods simultaneously so that the length of time and expense of the tests would be kept to a minimum. Figure 10 shows a cross-sectional diagram of the apparatus. Two knife edges (1) are fastened to stiff channels (2), and support a rod (3). A central knife edge (4) is attached to a container (5) holding sand. When the rod breaks, the container falls on the angles (7) which are supported on a different frame from the channels. In this way, when a rod breaks, the fall of the container does not create a shock that would seriously affect the other rods still on test. The apparatus is in three tiers, each simultaneously testing six rods. It is thus possible to conduct tests on 18 rods at a time. The apparatus

is located in the basement where the vibration conditions are quite satisfactory, as no heavy machinery or railroads are in the vicinity. Tests with dial indicators have shown no serious vibration present.

A table was constructed giving the load *versus* the extreme fiber stress for every rod diameter. In making a test, the diameter of the rod is measured, and the load is determined from the table. The container is then filled with sand to make up the proper load and a special levering device is used to apply the load gently. In general, an

accuracy of 1 percent could be attained in measuring the load and its duration. The 1,000-, 10,000-, and 86,400-second tests were made with the long-time loading device.

ACKNOWLEDGMENT

We are greatly indebted to Major J. M. McCormick, formerly of the Preston Laboratories, for his work and helpful suggestions in constructing the switching system whereby the labor in making tests on the short time loading apparatus was greatly reduced.

Fatigue of Glass under Static Loads

T. C. BAKER† AND F. W. PRESTON
Preston Laboratories, Butler, Pennsylvania
(Received November 1, 1945)

Static loading tests were run on glass and porcelain rods $\frac{3}{16}$ inch in diameter for times ranging in duration from 0.01 second to 24 hours, by using specially designed apparatus. It was found that glass can support for 0.01 second about 3 times the stress that would break it in 24 hours. The effects are generally the same for all glassy materials. Porcelain showed the effect somewhat less than glass. It appeared that adsorbed moisture and gases reduced the strength.

INTRODUCTION

IT has been recognized for many years that glass fatigues* under a perfectly steady load, breaking unexpectedly and without apparent cause under a stress it has long endured. Therefore, in measuring the mechanical strength of glass, the duration of the test must be specified as well as various other circumstances, some of which are discussed in this paper.

Whereas most glassware gets broken by impact, that is, by loads whose duration is measured in microseconds and which vary widely even in so short a time, yet some glassware is subjected to constant loads of long duration—for instance, soda, champagne, and beer bottles, aquarium windows, gauge glasses, and so on. An understanding of the relation between short time strength and long time strength is therefore important.

* This use of the term "fatigue" appears to us perfectly legitimate, though it is not the same as the usual engineering usage. If necessary, our fatigue may be distinguished as "static" fatigue, and the engineering usage as "dynamic" or "vibration" fatigue.

† Present address: Libbey-Owens-Ford Glass Company, Plaskon Division, Toledo, Ohio.

This relationship has already been worked out roughly over a substantial range of loading intervals by several experimenters, so that from a strictly practical point of view there is reason to believe we know enough to risk no hazards to the public. The interest, accordingly, shifts to the nature of "static" fatigue, and to such questions as whether the glass has any strength at all under an exceedingly long loading. If the strength after a great period is not zero, what is the value of it, and what is the meaning of it? On the other hand, we have reason to suspect that at loads of very short duration the strength may be very great. Is it theoretically infinite, or is it finite? If the latter, how great is it, and can any clear meaning be ascribed to it? Or is the question somewhat meaningless, in view of the inertia effects which make "steady" loads of very short duration inconceivable?

At first we had hoped our investigations would shed some light on these theoretical questions, but as more evidence accumulated all our beliefs as to the nature of fatigue in glass tended to change, and we felt much less inclined to specu-

late on the strength of glass and the ultimate effects of fatigue until the concept of strength was more precisely defined and the nature of fatigue was more thoroughly understood. Our initial feeling was that fatigue was due to the thermal effects at the tip of a flaw. The highly stressed bonds at that point were supposed to break when their momentary energies grew greater, due to the random shuffling of the thermal energies between the atoms. If this were the case, the cause of fatigue would be the same for long and short durations of load, and so we could expect a correlation between time and breaking strength. One might then assume that, if we had data encompassing a few orders of magnitude, we could extrapolate to both longer and shorter times with reasonable safety. However, as our work progressed it soon became apparent that atmospheric conditions were a factor in the fatigue of glass and that the phenomenon was so complicated that we could have little justification for extrapolation to either shorter or longer times, except as mere speculation.

The data presented are therefore disappointingly restricted, from the viewpoint of determining ultimate "strength" of glass. On the other hand, certain practical questions are answered as to the reliability of testing procedures, and at the same time a sounder understanding of the nature of the phenomenon has emerged, which may lead to a valuable line of research.

These tests were designed to investigate the phenomenon of strength decay, or fatigue, of glass in its normal condition, that is, with the amount of surface damage ordinarily present in commercial glasses, rather than the fatigue in glass having specially prepared, pristine surfaces. Glass samples were used, therefore, which had light surface damage easily visible under a microscope, which made their strengths comparable to that of commercial glassware. It is quite possible that all the breaks started at these flaws, and so the flaws in glass may be an essential part of the phenomena that we have studied. Other effects noticed in glass fibers, or other glass objects having pristine surfaces, may bear no simple relationship to the results obtained in the presence of large flaws.

Since it is not possible to experiment for an

indefinitely long time and report the results to the present generation, and since there are, on the other hand, practical difficulties in experiments for an infinitesimal time, we are limited to a range of duration which is not extremely extensive in "powers of ten" (orders of magnitude). From this relatively restricted range, we may perhaps hazard good guesses as to what might happen at somewhat shorter and somewhat longer periods of loading. The validity of our guesses depends on how accurately we can plot a curve of strength *versus* duration of load over the accessible range.

In the work herein reported, this range has been considerably extended beyond that of previous investigators. The extension is largely at the short duration end. This direction was chosen by one of us (T.C.B.) as offering more immediate promise than the other, because experimental conditions can be more readily kept constant for a brief time than for a long time. For instance, conditions such as temperature and humidity are important. Further, a lot of experimenting can be done in a matter of weeks or months at the short time end, while only a little can be done in many years at the other.

The essence of the short time loading experiments is the use of "electromagnetic loads," which can be very powerful while substantially free of inertia, so that the loads can be applied in 0.003 second without significant overtravel or oscillation. The quick-loading apparatus was described in a previous paper¹ in which its details and accuracy were discussed. This apparatus was not used for experiments lasting minutes or hours, since gravity provides a more convenient steady force. For these longer periods of testing, the essence of the set-up is the simultaneous testing of many specimens under conditions as constant as possible in respect to temperature and humidity.

The tests were made in flexure on $\frac{1}{8}$ -inch glass rods, the simplest method experimentally. The shortest practical duration was about 0.01 second with our apparatus. Theoretically, somewhat shorter times are practical in making tests in direct tension on fibers, but the experimental

¹ T. C. Baker and F. W. Preston, "Wide range static strength testing apparatus for glass rods," J. App. Phys. 17, 162 (1946).

difficulties were believed to be greater and apt to introduce more error. The longest tests lasted 24 hours. It is clear from the shape of the curves that little more would have been learned by extending the time to ten days, or even to one hundred or one thousand days. Furthermore, since each specimen must be tested at several successively greater loads, such tests would have taken many years to complete, even if experimental conditions could have been maintained constant.

It is not unthinkable that the curve may yet be extended, experimentally, farther than we have gone, but it seems at present unlikely that much practical, or even theoretical, information could be acquired thereby. Therefore, in the belief that the subject has reached a measure of completeness, we present an account of some of our experiments.

EXPERIMENTAL PROCEDURE

Because the strength of glass seems to be a function of its previous history of exposure both to the atmosphere and to mechanical contact, special precautions had to be taken in order to keep constant the effects of handling, adsorbed moisture, etc., in measuring the variation of breaking stress with changing duration of the load. Ideally, the controlled experimental procedure should start during the manufacture of the glass canes from which the rods were cut. Unfortunately, that was impossible. Our experimental procedure was constrained to begin after the glass reached the laboratory, which naturally left uncertain the previous history of the canes.

The initial step was to select canes that met certain specifications as to diameter and taper, and cut them into rods six inches long. These rods were then distributed among six or eight samples, of 25 rods each, and their diameters were measured and recorded. The rods were next set aside for a week or two so that any recent mechanical damage would tend to reach equilibrium, and the rods would have a chance also to come to equilibrium with the laboratory atmosphere. Finally, the rods were broken in the strength-testing apparatus.

Selection of Canes

In order to get comparable test samples, glass canes were used, whenever possible, which had

been made at the same time. After being cut into four-foot lengths the canes were sent to the Hartford-Empire Company to be annealed, where special precautions were taken to prevent extra damage during the annealing process. The annealed canes were then individually wrapped in tissue paper and sent to the Preston Laboratories. There each cane for a test lot was examined, and only those which met the following specifications were used.

Specifications for Cane

Taper per foot:	less than 0.005 inch
Out of roundness:	less than 0.005 inch
Diameter:	0.220 ± 0.010 inch
Radius of curvature of axis:	greater than 5 ft.
Canes showing unusual abrasion were rejected.	

Each cane was cut into six-inch lengths, or rods, and these were distributed among either six or eight samples, depending on whether the cane was three feet or four feet long. The method of distribution was as follows: the first rod from the first cane was put into the first sample, the second rod was put into the second sample, and so on. The second cane was distributed by putting the first rod into the second sample, the second rod into the third sample, and so on, until the last rod was put into the first sample. The subsequent canes followed out the progression. This method of distribution meant that each sample contained a rod from each cane, and each contained approximately the same number of rods cut from the ends and the middle of the canes. Thus 25 four-foot canes would make eight samples of 25 rods each, 25 three-foot canes would make six similar samples.

Care of Test Samples

As the canes were cut, the six-inch rods were put into wooden blocks having twenty-six numbered holes bored 1 inch deep. Only twenty of these specimens were used for the tests, the other rods being saved for spares. For example, if an accident occurred in the test on one rod, the rods taken from the same cane were discarded from all the other samples, and a complete set of spares was substituted. Great care was used in handling the rods during and after cutting, because new damage inflicted during subsequent handling could destroy the homogeneity of the specimens.

The ends of the rods were measured for maximum and minimum diameter, and a scratch was made at one end to indicate the direction of the minimum diameter of the rod. The central minimum diameter of the rod was estimated to be the average of the minimum diameters at both ends, and was recorded on the data sheet, next to the specimen number. In testing each individual rod the apparatus was adjusted to produce the appropriate stresses in accordance with special tables showing, for each value of load, the apparatus adjustment in terms of rod diameter.

The Apparatus

The rods were tested in cross bending, with a five-inch span. The load was applied by a single central knife edge. The two types of apparatus for testing the rods have been described in the previous paper. Briefly, one consisted of a modified dynamic loudspeaker to apply the loads of short duration (0.01 sec. to 100 sec.), and the other was a multiple loading device designed to accommodate numerous rods for the 1,000-sec. to 24-hour tests.

The load was applied to the rods as quickly as possible, left on for a predetermined time, and then removed as quickly as possible. The load was applied in two or three milliseconds and removed in $\frac{1}{2}$ millisecond in the 0.01-, 0.10-, and 1.0-second tests. In the 10- and 100-second tests about $\frac{1}{2}$ second was required to apply and to remove the load, and in the tests of 1000 seconds and over about four or five seconds were required. For the short time loading tests, tables were constructed giving the maximum stress in a rod as a function of the current through the armature of the dynamic loudspeaker, for all values of rod diameter. In making the tests, each rod was removed from the sample holder and placed in the testing apparatus with the minimum diameter in the direction of the load. This assured that the maximum fiber stress would be applied at the end of the minimum diameter. The rod was then tested at the minimum breaking load or at one loading interval below that. If the rod did not break, it was tested at an appreciably higher stress, the increment depending on the average strength.

In the case of the long time loading apparatus

the procedure was very similar, except that instead of the moving coil used in the electromagnetic loading a two-gallon can filled with sand was used to supply the load.

The Loading Increment

Glass "fatigues" under constant load, and unless it recovers promptly upon the removal of that load (a matter that is not certain), the breaking of the glass at a given load might be affected adversely by any previous loadings, even though these might be of less intensity. Experimental evidence had been accumulated in the past, however, that when step-wise loading is used, previous, lower loads had very little or no influence if they were 10 percent or so below the breaking load. Thus if our increments of loading are 10 percent or more above previous loadings, we can ignore the effects of the latter. For good measure, it might be well to make the increments of the order of 20 percent.

On the other hand, if we use large increments the statistical data will fall into very few class intervals and the average will be uncertain. However, it was determined that, when we had five class intervals, the error would probably be quite a bit less than that from other sources. Accordingly, the loading increment chosen was 500 lb./sq. in. for samples whose averages were between 3000 and 6000 lb./sq. in.; 1000 lb./sq. in. for samples whose averages were between 6000 and 12,000 lb./sq. in.; and 2,000 lb./sq. in. for samples whose averages were over 12,000 lb./sq. in. The strength was generally found to vary over a 2 to 1 range for any sample. Thus in a sample with 12,000 lb./sq. in. average strength some specimens would break at about 8000 lb./sq. in. and some at about 16,000 lb./sq. in. This resulted in five class intervals for the data, using steps of 2000 lb./sq. in.

If a rod survived a test at 14,000 lb./sq. in. and broke at 16,000 lb./sq. in. it would be recorded as having broken at 16,000 lb./sq. in. Obviously, it could have broken at a load anywhere between 14,000 and 16,000 lb./sq. in., and on the average would break at 15,000 lb./sq. in. Therefore, it was necessary to apply the usual statistical correction for the magnitude of the class interval. When the loading increment was 2000 lb./sq. in.

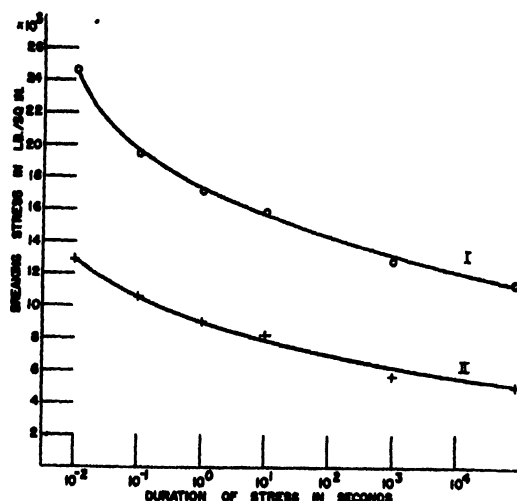


FIG. 1. I. Disannealed Pyrex, dry; II. Disannealed Pyrex, scratched, wet.

the correction would be -1000 lb./sq. in. When the loading increment was $1,000$ lb./sq. in., then the correction would be -500 lb./sq. in.

Special Precautions

Special precautions were necessary to keep the temperature and atmospheric conditions as constant as possible during the tests. The room temperature was kept as near 75°F as possible. The rods were placed near the apparatus previous to the test so that thermal strains would not be introduced when the rods were inserted in the apparatus. A few minutes before the test the rod was wet with water at room temperature, and water was applied to the rod whenever it became dry during the successive loadings.

RESULTS OF VARIOUS STATIC LOADING TESTS

The results herein reported were obtained from tests run on five glasses and three porcelains. These are presented in Table I, which gives, for each loading period, the average breaking strength of 20 to 25 specimens. The results for the various glasses are plotted in Figs. 1 and 2, and the results for the porcelains are plotted in Figs. 3 and 4. All results are corrected for the loading increments, as described in the preceding section.

Lot I consisted of Pyrex rods which had already been cut to length, since there were no long canes available at the time of the tests. In that respect, this constituted an exception to our usual procedure. The rods were rather strongly disannealed, with a surface compression of 2000 to 4000 lb./sq. in., the average of several rods being about 3000 lb./sq. in. Moreover, as these were our first tests and the effect of moisture was not appreciated at that time, the Pyrex rods were tested dry. Various difficulties, including the fact that no really reliable average for the surface compression could be taken, greatly reduced the accuracy of these results.

Lot II was made up of the same type of disannealed Pyrex rods as used in lot I except that they were scratched circumferentially with a medium grade of emery. By this time the effects of humidity had been detected and the rods were tested wet.

Lot III consisted of samples made from four-foot, annealed, machine-drawn, soda-lime canes which were unusually straight and round and made an excellent stock for selection of samples.

TABLE I. Average breaking stress in pounds per square inch for various durations of load.

Lot	Material	Test Wet or Dry	0.01 sec.	0.1 sec.	1.0 sec.	10 sec.	100 sec.	1000 sec.	86,400 sec.	Ratio of 0.1-sec. to 24-hour test	Ratio of 0.01-sec. to 24-hour test
I	Disannealed Pyrex	Dry	24,600	19,600	17,100	15,900		12,800	11,300	1.7	2.2
II	Disannealed scratched Pyrex	Wet	12,900	10,600	9000	8250		5750	5050	2.1	2.5
III	Annealed soda-lime glass	Wet	19,900	15,900		10,500		7950	6450	2.4	3.1
IV	Annealed lead glass	Wet		19,000	14,900	11,200		9400	7950	2.4	
V	Fused silica	Wet		23,880	12,300	19,800	15,660	14,280	11,740	2.0	
VI	Porcelain Type A	Dry	15,800	13,900			9700		9100	1.53	1.74
VII	Porcelain Type B	Dry	15,400	13,500	11,900	9900	9400	9100	8900	1.52	1.73
VIII	Porcelain Type C	Dry	10,900	9400	8400	6900		6000	5900	1.59	1.85

Their composition was very close to that of bottle glass. Considerable care was taken in annealing and preparing comparable samples, and when certain check tests were run, the average strengths agreed to better than one percent, indicating excellent uniformity of the samples. These rods were tested wet.

Lot IV was composed of glass rods containing approximately 50 percent lead by weight. The glass was unusually soft and easy to scratch, and the resulting damage on some rods made the tests somewhat erratic. These rods were annealed, and the tests were conducted with the rods wet.

Lot V consisted of rods made from silica-glass canes three feet long. As the rods were apparently free of stress, no special annealing treatment was given them. All rods were tested wet.

Lot VI was composed of porcelain rods* of the following composition, fired at cone No. 9.

22 percent Kentucky Ball Clay
25 percent Florida Kaolin
37 percent Canadian Feldspar
10 percent Pennsylvania Flint
6 percent Water

The rods were tested dry because the effect of

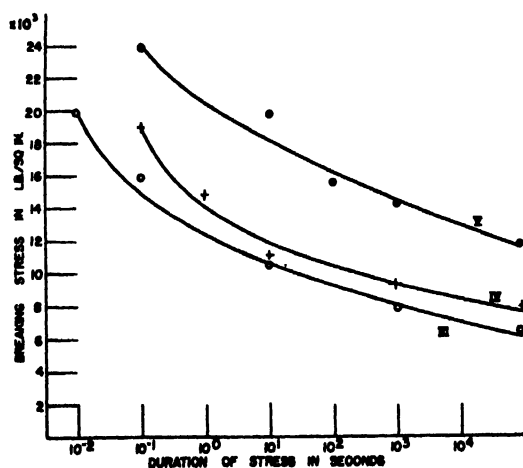


FIG. 2. III. Annealed soda-lime glass, wet; IV. Annealed lead glass, wet; V. Fused silica, wet.

moisture on the strength of porcelain is considerably less than on the strength of glass. The possible errors caused by not testing wet are probably less than two or three percent.

* Porcelain rods used in these experiments were supplied by George Aderhold of the Saxonburg Potteries, Saxonburg, Pennsylvania.

Lot VII had the same composition as lot VI but the porcelain was fired at cone No. 10 instead of cone No. 9. These rods were also tested dry.

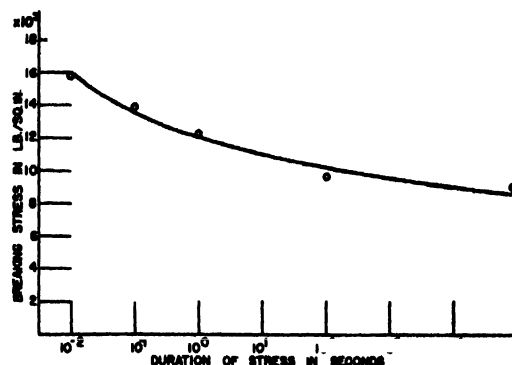


FIG. 3. VI. Porcelain, type A, dry.

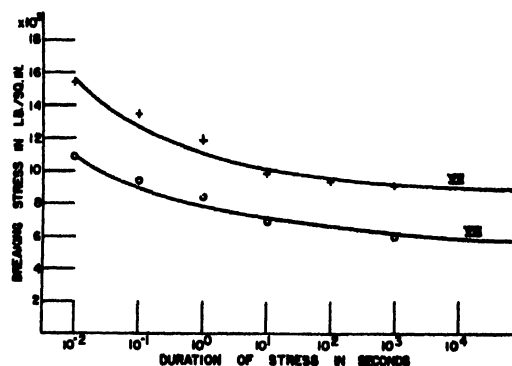


FIG. 4. VII. Porcelain, type B, dry; VIII. Porcelain, type C, dry.

Lot VIII consisted of porcelain rods of the following composition, fired at cone No. 12:

50 percent Canadian Feldspar
10 percent Florida Kaolin
30 percent Kentucky Ball Clay
10 percent Water

The rods were tested dry.

Of the results shown in Table I, those for soda-lime glass, lead glass, and fused silica may be considered most reliable. In the case of the Pyrex glass the records are ambiguous regarding the amount of the correction that was applied for the disannealing present in the rods. The amount of surface stress in the porcelain rods due to disannealing could not be determined, so some error from this cause may be present in these

data also. The rods of lots III-V, however, were known to be free of any disannealing stress and these data, therefore, do not include the possibility of such an error.

DISCUSSION OF ERRORS

At the present time, an exhaustive treatment of the errors in the fatigue testing of glass is quite impossible, because of the entanglement of the variables, which cannot as yet be independently investigated. However, a partial list of the factors will be given so that the reader may be assisted in evaluating the results.

1. Errors in Apparatus

In a previous paper the errors involved in applying the load and the errors of timing the duration of load were discussed. In general it was felt that these errors were not greater than ± 5 percent.

2. Errors in Measuring the Rod Diameters

This error was not large because the diameters could be measured to 0.0005 inch, or $\frac{1}{4}$ percent. Such an error would lead to only $\frac{1}{4}$ percent error in the applied stress, and since it was a statistical error it would practically disappear in the average for twenty rods.

3. Errors Due to Single Knife Edge

Since the rods were measured in cross bending, with a 5-inch span and single knife edge loading, a certain amount of error existed when the rod did not break exactly underneath the knife edge. At a quarter-inch away from the central loading point the stress would be 10 percent less than at the center. This changing stress caused a peculiar type of error to enter the results, because the long time loads nearly always broke the rod exactly under the knife edge, whereas under the short time loads the rod often broke as much as $\frac{1}{4}$ inch away from the central point. This would lead to a correction upwards in the short time strength, because the portion under highest stress was not broken, and therefore would have broken only at a still higher stress. Scratched rods nearly always broke under the knife edge, even in short time tests. The effect seems also to be a

function of the number and kind of flaws. It is estimated that this error could reduce the average by as much as 5 percent in the short time loads.

4. Error Due to Scattering of the Data

The strength of the strongest rod in a sample might be twice as great as the weakest one. This scattering of the data resulted in a standard deviation for the individual rods that was about 10 percent of the average of the sample. The standard deviation of the average of 20 to 25 rods would be about 2 to 2.5 percent of the average. It would appear from the foregoing that the average strengths of different samples could be expected to agree to only 5 percent, and differences of as much as 8 percent would not be infrequent. On the other hand, actual check tests indicated that the procedure of preparing the samples had led to a strong correlation between them and that generally the average strength from sample to sample agreed to one percent or better. Unfortunately, insufficiency of data prevents an accurate computation of the correlation coefficient; and there is no assurance that the averages of 20 to 25 rods might not differ by as much as 5 percent due to purely statistical fluctuations.

5. Errors Due to the Loading Increments

The choice of the loading increments involved a compromise between two errors. One error would be that caused by the weakening of the rod from previous loading, and would become greater as the increment was made smaller. The other error would be due to an insufficient number of statistical class intervals, and would increase as the increments were made larger. This latter error is fully described in books on statistics² so that it need not be discussed here, except to state that it would probably be less than one percent in the average, when a loading increment of between 10 and 20 percent of the average is used.

The weakening effect of the previous loading has not been fully investigated. Some tests were made which indicated that a loading increment

² See, for example: G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics* (Charles Griffin and Company, Ltd., London, 1940) Chap. 6.

of 10 percent of the average or more, would cause less than a one percent decrease in the average strength.

6. Errors Due to the Atmospheric Conditions

Considerable data were obtained indicating that moisture and possibly gases affect the strength of glass. The specimens were tested wet after the phenomenon was discovered, but even then the presence of gases such as CO₂ possibly affected the results from one to five percent, the greater effect being for the longer load durations.

7. Errors Due to Temperature

Evidence has been obtained showing that temperature has a profound effect on both the strength and the fatigue of glass. It turns out that even small changes such as occur indoors will cause a change in the strength of glass amounting to one or two percent.

8. Errors Due to Storage of Rods

Since it was impossible to make all the tests at the same time, some of the rods were stored longer than others while the atmospheric conditions changed. There was some evidence that when rods were stored in a dry atmosphere for some time they did not have the same strength as rods stored in a damp atmosphere, even though they were tested wet in both cases. This resulted in another possible error of one or two percent in the final averages.

9. Combination of Errors

Considering these errors as a whole, it is estimated that the results cannot be considered accurate and controlled to more than ± 6 percent. Moreover, it must be appreciated that certain types of error will tend to distort or bias the data, making the short time strengths appear less than they should be in comparison with the long time strengths, or *vice versa*. Therefore, any generalizations that depend upon an accuracy of better than 10 percent in the results, or series of results, should be treated with extreme caution.

DISCUSSION OF RESULTS

Since a primary objective of this work was to investigate the fatigue in commercial glassware

under normal conditions of temperature and humidity, the data are limited to normally scratched glass tested either wet or in an atmosphere of high humidity. As the difference in the strength of glass between 30 percent humidity and completely wet is only about 5 percent, the results are quite comparable. The surface condition of the rods was similar to the normally scratched and abraded condition found on bottles and window glass. Scratches, chatter marks, and percussion cones could be easily seen under a 50-power microscope.

Amount of Fatigue in Glassy Materials

The fact most apparent from the data is that all of the glassy materials tested exhibit the phenomenon of static fatigue to a marked degree. In the case of the soda-lime rods the 0.01-second strength is 3.1 times the 24-hour strength. The ratio of the 0.1-second strength to the 24-hour strength is between 1.7 and 2.4 for all the glasses tested. The corresponding ratios for porcelain ranged from 1.52 to 1.59.

Nature of Fatigue Curves

The plotted curves giving the relationship between the breaking strength and the logarithm of the duration of test show that there is a rapid decrease in strength up to a minute or so, and a very slow decrease from then on. The relationship between the strength and the logarithm of the duration of load is not linear, as has been previously supposed.³ At present there is some question as to whether or not the curve is truly a monotonic decay, with steadily decreasing slope and steadily increasing radius of curvature, as we go to longer times. There remains a possibility, within the limit of accuracy of the tests, that there may be several points of inflection, implying deviations from our drawn curves.

Effect of the Condition of the Surface upon Fatigue

The condition of the surface has a profound effect on fatigue as well as upon the strength of the glass. Scratched glass loses much less strength in a given time interval than does unscratched glass. However, the percent loss of strength is roughly the same. This may be seen by comparing

³A. Smekal, *Ergebr. d. exakt. Naturwiss.* 15, 106-188 (1936).

the tests on the scratched and unscratched Pyrex rods. The decrease for the scratched rods was 7850 lb./sq. in. and the decrease for the unscratched rods was 13,300. Yet the ratios of the 0.01-second strength to the 24-hour strength were 2.5 and 2.2, respectively. Unfortunately, this is all the data that are available for the comparison of scratched with unscratched rods, and the variable annealing stresses that existed in these rods as well as the different moisture conditions, tended to obscure the results.

That badly damaged glass fatigued in the same ratio as less damaged glass was borne out by the fact that the percent standard deviation of a sample of each kind tested at 0.01 second was approximately the same as that of a sample tested for 24 hours, although the strengths had changed by a factor of 3.1 to 1. Thus in the case of the soda-lime glass the standard deviation for the 0.01-second test, in which the average was 8500 lb./sq. in., was 800 lb./sq. in. In both cases the Pearson coefficient of variation was approximately 10 percent. Thus the stronger rods lost more strength than the weaker ones, yet they lost it in the same ratio. On the other hand, the fact that the rods broke under the central knife edge in long duration tests, but tended to break at some distance from it during the short duration tests, indicates that strong glass (i.e., glass with little surface damage) has a greater percentage loss of strength than weak glass. For it may be argued that in the short time tests some flaws made the rods weaker than other flaws did and the rods would break at 10 percent less stress. If these flaws were only $\frac{1}{4}$ inch away from the central knife edge, the rod would break there. Then if the minor and major flaws reduced the strength in the same ratio, we should get the breaks in about the same way on the long time tests. The fact that this does not happen indicates that the minor flaws must cause a greater fatigue rate in order to originate the breakage on the long time tests.

Effect of Chemical Composition

Because the loss of strength of glass is connected with surface damage, which cannot be

accurately standardized between different glasses, it is difficult to make accurate judgments concerning the effect of chemical composition upon fatigue. Apparently the short time to long time strength ratios are about the same for all glasses studied. The results on the disannealed Pyrex rods have to be discounted because of the impossibility of applying a good correction for annealing stresses. We are left, therefore, with the results for silica glass, lead glass, and soda-lime glass as being perhaps accurate enough to judge this point. The ratio between the 0.10-second test and the 24-hour test is 2.4 in the last two cases, so no distinguishable difference is evident. The ratio however is only 2.0 in the case of the silica glass, so there is some evidence that silica glass fatigues slightly less than soda-lime or lead glasses.

In the case of the porcelains there is no question that they fatigue much less than the glass, as the ratio of the 0.10-second test to the 24-hour test ranges from 1.52 to 1.59. However, so little is known about the nature of fracture in porcelains, or aggregates in general, that it would be futile at this time to try to explain the fatigue in porcelain in terms of the fatigue phenomenon in glass, although they are probably closely related.

SUMMARY OF RESULTS

The results have corroborated previous investigations showing that the strength of glass depends on its surface damage, on the time it is under load, and upon atmospheric conditions. As the work progressed it became evident that these factors were interdependent in a complicated way. Nevertheless, to an accuracy of about 10 percent we can be assured that glass has about three times the strength under our briefest loads that it has under long time loads; that the relative loss of strength with time is about the same for lightly damaged and badly damaged glass surfaces, although the absolute strength varies radically with the damage; and finally, that chemical composition has only a small effect, if any, on fatigue.

The Effect of Water on the Strength of Glass

T. C. BAKER† AND F. W. PRESTON
Preston Laboratories, Butler, Pennsylvania

(Received November 1, 1945)

The strength of glass was measured in various surrounding media. It was found that glass is 20 percent stronger when dry than when wet, and two to two and a half times as strong when baked in vacuum than when tested wet (on tests of ten seconds duration). Apparently, the chief cause of the loss of strength, as compared with that in vacuum, is moisture, but gases, especially CO₂, seem to have some effect. It was found that fatigue of glass disappears when the glass is tested in a vacuum.

Autoclaving tests showed that glass is rapidly etched by water at higher temperatures, and that even silica glass is considerably attacked. It appears that the attack of water on glassy surfaces produces gels, and this attack goes on inside the flaws, thus considerably complicating the problem of the strength of glass.

INTRODUCTION

THE previous two papers^{1, 2} have shown that, under the conditions of the experiments therein described, there is a precipitate fall in the breaking strength of glass, including fused silica, and of porcelains in the first few milliseconds of loading, and a slower but steady fall with the lapse of further time. They have also shown that moisture seems to have something to do with this, so that the breakage of glass can perhaps not be understood simply in terms of stress, strain, and elastic constants; rather does "strength," as we usually measure it, involve the properties of the surrounding medium, perhaps a reaction with the surrounding medium, and so pass out of the field of engineering and even of physics into the domain of physical chemistry.

There have, indeed, both prior to the experiment detailed below and also between the time of the experiments (c1937-39) and the present (1945), appeared several papers by other investigators on the effect of the surrounding medium. Milligan,³ for instance, found that exposing freshly cracked glass to water vapor or alcohol vapor considerably reduced its strength. He found that moistening a freshly scratched piece

of glass reduced its strength 20 percent or more, and immersing the freshly scratched glass in dry paraffin oil increased the strength about 20 percent. Initial experiments of the authors indicated that changes of the relative humidity of the atmosphere would affect the results of strength measurements by 5 percent or more. It was thus evident that the surrounding medium played a very important part in tests for the strength of glass.

So it appeared desirable to investigate the effects of the surrounding medium on the strength of glass in greater detail than had been done heretofore. It was hoped that the studies would shed light on the causes of fatigue and possibly lead to a more complete understanding of the phenomenon of breakage.

Accordingly, the glass specimens were tested after they had been put into various solutions, and after they had been dried in desiccators; they were even tested in vacuum. Unfortunately, the results of the tests were generally not clear cut, and often it was not possible to duplicate them, so the general picture became more complicated with every new experiment. At the time the work was stopped a great many loose ends remained, and our understanding of the strength of glass had not become sufficiently fixed so that we could feel sure of our ground. Nevertheless, from a study of our data, in combination with that available in the literature, a general picture of the strength of glass began to emerge, indicating a much greater complexity than is usually supposed.

† Present address: Libbey-Owens-Ford Glass Company, Plaskon Division, Toledo, Ohio.

¹ T. C. Baker and F. W. Preston, "Wide range static strength testing apparatus for glass rods," *J. App. Phys.* 17, 162 (1946).

² T. C. Baker and F. W. Preston, "Fatigue of glass under static loads," *J. App. Phys.* 17, 170 (1946).

³ L. H. Milligan, "Strength of glass containing cracks," *J. Soc. Glass Tech.* 13, 351 (1929).

First of all, commercial glass has tiny flaws in its surface which can easily be seen with a microscope. After the flaw is made, water and gases enter it and radically change the nature of the surfaces and alter the physical and chemical conditions at the tip of the crack. If the glass at the flaw is under tensile stress, the water vapor affects the bonds at the tip of the flaw, causing great weakening. If the flaw is allowed to stand, the water vapor forms thin gels on the surfaces and rounds out the tip of the flaw, causing an increase in strength. There is reason to believe that complete equilibrium was not established in any of our experiments.

The "strength" of glass began to appear to have little meaning as a physical constant of the material. It might be more correct to say that the phenomenon of rupture is due to several causes, including tensile stresses at the surface of the glass, the time during which the tension is applied, the nature and depth of the flaws, and the penetration of atmospheric or other reagents into the inside of the flaws. It is obvious that a great stress in itself would be a sufficient condition to break the glass, yet there exists evidence that atmospheric and weather conditions also would cause some flaws to extend to the point of rupture.⁴ So it is doubtful that any one single cause is an essential element in the breakage.

With such a complicated picture to investigate it becomes apparent that controlled tests are particularly difficult. Many of the factors cannot be truly controlled, such as the size and orientation of a flaw, so it is not surprising that effects should occur that could not be duplicated. For this reason much of the data in this paper has been included not to prove points finally but merely to indicate the nature of effects, so that future research on this problem will be on a sounder basis.

The strength of ordinary commercial glass depends primarily on the size, depth, and orientation of the tiny cracks and flaws on its surface. These flaws cause the strength to be reduced from a theoretical strength of several million pounds per square inch to one that varies between 2000 lb./sq. in. and 20,000 lb./sq. in. in

glassware whose surface has been damaged by abrasion and impact. The effects of the surrounding medium upon the strength, therefore, can only be comparative. So, in our experiments, two or more samples of about 25 glass rods each were tested under different conditions and the ratio of their strengths determined.

Any surrounding medium seems to lower the strength by a certain ratio below the vacuum strength, rather than to subtract a fixed amount from the vacuum strength. For example, if the strengths of two comparable samples were obtained, one in vacuum and the other in air, and the strength in vacuum was 20,000 lb./sq. in. and in dry air 10,000 lb./sq. in., the ratio would be 2.0. Two other somewhat stronger samples tested in vacuum and dry air might have strengths of 30,000 and 15,000 lb./sq. in., and in this case also the ratio would be 2.0. This rule, that the surrounding medium affects the strength by a ratio, seems to be approximately true of all our studies on fatigue of glass as affected by the surrounding medium.

APPARATUS

The apparatus used to measure strengths of glass rods in the following experiments was the quick-loading apparatus described in a previous paper.¹ When it became necessary to measure the strength of glass in vacuum, the quick-loading apparatus was put into a vacuum chamber. Otherwise the rods were tested out in the open air after they had been treated according to the various procedures described.

The vacuum strength-testing apparatus included a vacuum chamber, consisting of a bell jar resting on a cylindrical steel shell, in which the entire quick-loading apparatus was placed. A glass rod is supported by two knife edges five inches apart. A wire bale is looped over the glass rod and connects to the armature which hangs freely in a magnetic air gap. A field coil produces the magnetic field in the air gap. This field coil was covered with cement to prevent outgassing when the vacuum was drawn. On either side of the glass rod were added two heater coils consisting of chromel wire wrapped around porcelain rods. An easily removable furnace shield prevented radiation from the heater coils from

⁴See A. Silverman and L. G. Ghering, "Deterioration of whiskey bottles," *Glass Ind.* [9] 15, 339 (1942).

warming the bell jar too much while the rod was being baked out. A glass window at the end of the furnace shield permitted observation of the rod during the test. The shield had several openings to the main chamber so that there was no doubt of its having the same vacuum as that in the main chamber. The vacuum chamber is enclosed by a large bell jar resting on a steel cylinder which has a flat bottom welded to it. The vacuum chamber is evacuated through an outlet leading to a McLeod gage and a diffusion pump.

PROCEDURE

The procedure for testing the strength of glass rods was described in a previous paper.²

Briefly, a table was constructed which gave the stress in terms of the current through the armature coil for every rod diameter. The diameter of each rod was measured before inserting it into the quick-loading apparatus. A minimum stress was put on the rod for a given length of time; if the rod did not break, it was then stressed at a higher level for the same length of time, and so on until it broke. The stress was increased in steps of 2000 lb./sq. in. if the average strength was greater than 12,000 lb./sq. in., and in steps of 1000 lb./sq. in. if the average strength was less than 12,000 lb./sq. in. Preliminary experiments enabled the operator to tell which increment to use. The average breaking stress of each sample of rods was corrected, as explained previously,² by subtracting one-half the increment from the average.

The procedure in testing glass rods for strength under vacuum was to insert the rod in the quick-loading apparatus, put the furnace shield into position and, after placing the bell jar on top of the iron cylinder, start the mechanical vacuum pump. Because of the large volume of the test chamber, the mechanical pump was operated for a half-hour or so before it became feasible to start the diffusion pump. After an hour and a half the vacuum would be down to 1 micron or less, at which time the rod would be baked out at about 350°C for 10 minutes. After baking out, the rod was allowed to cool for one hour because it lost heat very slowly. Even at the end of an hour, when the rod was tested for strength, the temperature of the rod was about 80°C.

TABLE I.

Test No.	Pressure	Temp. of bake out	Time of bake out	Duration of test	Average strength
1	0.05 micron Hg	350°C	10 min.	10 sec.	24,700 lb./sq. in.
2	0.2 micron Hg	350°C	10 min.	10 sec.	23,800 lb./sq. in.
3	0.1 mm Hg	350°C	10 min.	10 sec.	18,900 lb./sq. in.
4	760 mm (wet)			10 sec.	12,100 lb./sq. in.

RESULTS

Results of the Vacuum, Drying, and Autoclaving Tests

The strength of glass rods was measured under various degrees of vacuum with different heating treatments. These tests were aimed at discovering the effect of vacuums, and the effect of heating plus vacuums, upon the strength and fatigue of glass of various chemical compositions. In general, comparisons were made between the average strengths of twenty rods. In this way a series of twelve experiments was carried out on comparable samples to determine the effect of various treatments on "strength."

Experiment I—Variation of Vacuum, Soda-Lime Glass

Soda-lime rods approximately two years old were made into four comparable samples of approximately twenty rods each. The rods had been stored part of the time in an unheated warehouse and had been alternately moist and dry, causing a slight amount of weathering or bloom to form on their surfaces. Three of the samples were tested under vacuum as described above. The fourth sample was tested wet under atmospheric pressure with no heating. The results are shown in Table I.

From these results it can be seen that the strength is greatest in the highest vacuum. The strength of the glass when tested in the 0.05-micron vacuum is greater than the strength when the glass is tested wet at atmospheric pressure by a factor of 2.04. The tests are limited to ten-second duration. It is probably significant that the strength at 0.05 micron is slightly greater than at 0.2 micron, but since the probable error of the results is around 250 lb./sq. in. from statistical fluctuations alone, this conclusion will have to be treated with reserve.

TABLE II.

Test No.	Pressure	Temp. of bake out	Time of bake out	Duration of test	Average strength
1	0.1 micron Hg	350°C	10 min.	10 sec.	44,300 lb./sq. in.
2	0.1 mm Hg	350°C	10 min.	10 sec.	32,200 lb./sq. in.
3	10 mm Hg	350°C	10 min.	10 sec.	29,200 lb./sq. in.
4	760 mm Hg (wet)			10 sec.	17,700 lb./sq. in.

Experiment II—Variation of Vacuum, Pyrex Glass

Four comparable samples of carefully handled Pyrex rods were tested by the same procedure as used for the soda-lime rods in experiment I. The results are shown in Table II.

The strength is greatest in the highest vacuum and is less as the vacuum is reduced. In this experiment on Pyrex glass the strength in the 0.1-micron vacuum is 2.50 times the strength wet in the atmosphere.

Experiment III—Silica Glass

This experiment was made on two comparable samples of silica glass rods. The procedure was the same as above. The results are shown in Table III.

These tests show that for silica glass the ratio between the strength of rods baked out in vacuum and the strength of those tested wet is only 1.67 on the ten-second test.

TABLE III.

Test No.	Pressure	Temp. of bake out	Time of bake out	Duration of test	Average strength
1	0.05 micron Hg	350°C	10 min.	10 sec.	27,900 lb./sq. in.
2	760 mm Hg (wet)	350°C	10 min.	10 sec.	16,700 lb./sq. in.

Experiment IV—Porcelain

In order to see whether or not a similar effect is obtained with porcelain, strength tests were made on three comparable porcelain samples. For the first sample the procedure was to bake out the test specimens in a vacuum for 10 minutes and allow them to cool for an hour to about 30°C before testing. On the second sample the rods were dried for one week in a desiccator in the presence of phosphorous pentoxide, then taken out of the desiccator and quickly tested in the test chamber at atmospheric pressure in

the presence of fresh phosphorous pentoxide. The third sample was tested wet, at atmospheric pressure. The results are shown in Table IV.

From these tests it can be seen that the ratio between the strength in vacuum and the strength wet is only about 1.5. The effect is apparently markedly less than on glass rods.

Experiment V—Effect of Heat, Soda-Lime Glass

There was some question as to whether or not heating had much effect upon the strength of glass rods under vacuum. This test was made to compare the strength of glass rods under vacuum

TABLE IV.

Test No.	Atmosphere	Temp. of bake out	Time of bake out	Duration of test	Average strength
1	0.01 micron vacuum	350°C	10 min.	10 sec.	10,400 lb./sq. in.
2	Dried in desiccator			10 sec.	7400 lb./sq. in.
3	Wet			10 sec.	6950 lb./sq. in.

with and without baking out. Since from the time the vacuum pumps were started to the time of test on the baked-out rods was about four hours, the sample that was not baked out was kept under vacuum for that time also. The results (Table V) were obtained on three comparable samples of 20 soda-lime rods.

These tests clearly showed that the strength was greatest on the 10-second tests when the rods were baked out in vacuum.

Experiment VI—Effect of Vacuum on Fatigue, Soda-Lime Glass

In order to study the effect on fatigue of different levels of vacuum, four comparable samples of soda-lime glass rods similar to those in experiment I were used. These rods had been stored in an unheated warehouse and had visible

TABLE V.

Test No.	Pressure	Temp. of bake out	Time of bake out	Duration of test	Average strength
1	0.1 micron Hg	350°C	10 min.	10 sec.	25,950 lb./sq. in.
2	0.1 micron Hg	None		10 sec.	21,900 lb./sq. in.
3	760 mm Hg (wet)			10 sec.	12,900 lb./sq. in.

evidence of weathering on their surfaces. The results are shown in Table VI.

These results indicate that there is no fatigue when the rods are baked out at 0.1 micron of Hg vacuum, but there is considerable fatigue at 0.1 mm of Hg vacuum. The fact that the strength of the rods on the 0.1-second tests was greater for the lower vacuum than for the higher vacuum seems hard to explain, especially since the results could not be duplicated in experiment VII which was performed on another batch of soda-lime rods of the same chemical composition. However, when Test No. 3 was being conducted, it became evident as soon as a few rods had been tested that the data were going to be very different from that of Test No. 1, and the apparatus was carefully checked without discovering anything unusual. There was no doubt of the increased stress the rods were bearing; this was shown by the amount of bending they underwent before breaking. Moreover, check

TABLE VI.

Test No.	Pressure	Temp. of bake out	Time of bake out	Duration of test	Average strength
1	0.1 micron Hg	300°C	10 min.	0.1 sec.	25,900 lb./sq. in.
2	0.1 micron Hg	300°C	10 min.	10 sec.	25,900 lb./sq. in.
3	0.1 mm Hg	300°C	10 min.	0.1 sec.	32,200 lb./sq. in.
4	0.1 mm Hg	300°C	10 min.	10 sec.	21,600 lb./sq. in.

tests with a few remaining canes were run at conditions of Test 3, which confirmed the above results on this batch of rods.

Experiment VII—Effects of Vacuum and Heating on Fatigue

In order to investigate the peculiarities noticed in experiment VI, Tests 1 and 3, a series of tests was made on a new batch of soda-lime rods of the same chemical composition as those used in experiment VI. These rods were only a few months old and showed no signs of weathering at the time the tests were made. Various heat treatments were used to see if there was any indication of the effects found in experiment VI. The results are shown in Table VII.

These tests confirm the results of experiment VI in that there is practically no fatigue when the glass is baked out at high vacuum. The tests do show however that the increased strength

TABLE VII.

Test No.	Nature of treatment	Pressure	Duration of test	Average strength
1	Heated 15 minutes at 0.01 micron using P_2O_5	0.01 micron	10 sec.	25,520 lb./sq. in.
2	Heated 15 minutes at 0.01 micron, no desiccator	0.01 micron	10 sec.	25,330 lb./sq. in.
3	Heated 15 minutes at 0.1 mm; then again at 0.01 micron	0.01 micron	0.1 sec.	25,809 lb./sq. in.
4	Heated 15 minutes at 0.1 mm, tested cool	0.1 mm	0.1 sec.	25,619 lb./sq. in.
5	Heated 5 minutes at 0.1 mm	0.1 mm	0.1 sec.	25,050 lb./sq. in.
6	Heated 10 minutes at 0.1 mm, held 1 hr.; heated 19 minutes	0.1 mm	0.1 sec.	23,050 lb./sq. in.
7	No heating	0.1 mm	0.1 sec.	25,250 lb./sq. in.

on Test 3 of experiment VI was not a repeatable phenomenon on the newer rods, and that the 0.1-second strength at lower vacuum was practically the same as that at higher vacuum. It is our opinion that weathering had formed gels in the cracks that had very much influenced the strength tests in experiment VI.

Experiment VIII—Effect of Dried Air

In order to get some idea as to whether or not gases other than water vapor affect the strength of glass, four comparable samples of soda-lime rods were prepared. In the first test the rods of sample 1 were baked out and tested in vacuum as before. In the second test the rods of sample 2 were baked out at 0.01 micron and allowed to cool for 1 hour, then the air was allowed to enter after it had been dried by bubbling through two stages of concentrated sulphuric acid. After a period of half an hour, the rods were tested for strength. The third test, with sample 3, was the

TABLE VIII.

Test No.	Pressure	Temp. of bake out	Pressure of test	Duration of test	Average strength
1	0.01 micron Hg	300°C	.01 micron	10 sec.	25,500 lb./sq. in.
2	0.01 micron Hg	300°C	760 mm after $\frac{1}{2}$ hour	10 sec.	22,900 lb./sq. in.
3	0.01 micron Hg	300°C	760 mm after 24 hours	10 sec.	22,000 lb./sq. in.
4			760 mm wet	10 sec.	12,500 lb./sq. in.

TABLE IX.

Test No.	Relative humidity	Duration of test	Average strength
1	0	100 sec.	11,300 lb./sq. in.
2	30%	100 sec.	10,200 lb./sq. in.
3	100%	100 sec.	9,800 lb./sq. in.

same as the second except that the rods remained in the dry air atmosphere for 24 hours before testing. The rods of sample 4 received no heat treatment and were tested wet at atmospheric pressure. The results are shown in Table VIII.

Apparently even dry air causes some reduction in the strength of glass, as can be seen from a comparison of Test 1 with Tests 2 and 3. However, it is possible that slight amounts of moisture were able to come through the drying apparatus and of course there must have been some SO_2 and H_2SO_4 gases present.

Experiment IX—Effects of Dry Storage

It was thought advisable to obtain comparative figures on the strength of "dry" glass rods as compared with glass under average atmospheric conditions and as compared also with thoroughly wet glass. The "dry" rods were dried in a desiccator for a week in the presence of P_2O_5 before testing. The results are shown in Table IX.

An increase of strength of about 15 percent above the strength when wet is indicated when rods are dried in a desiccator, and an increase of about 5 percent when the relative humidity is around 30 percent. However, this work was not too carefully controlled and is included merely to show approximately what can be expected.

Experiment X—Effect of CO_2 , Soda-Lime Glass

A few tests had indicated that some factor other than water was a weakening agent. We suspected CO_2 because it is known that CO_2 in weak solution will dissolve silica more than will plain water. Four comparable samples were used. Each sample was put into a different liquid for two hours previous to testing and kept wet with the liquid during test. The liquids used were distilled water, distilled water carbonated at room temperature, distilled water with satu-

rated $\text{Ca}(\text{OH})_2$, and distilled water with normal NaOH . It was felt at this time that the differences of strength observed were due more to differences of fatigue rates rather than to a fundamental lowering of the "strength." Therefore the rods were tested for 100 seconds duration in hopes that differences would be more apparent. The results are shown in Table X.

There is some indication that the CO_2 is an additional weakening agent besides the water. If so, the rest of the data might be explained on the basis that the CO_2 that is normally in the air was kept away from the flaws. However, until more carefully controlled tests are available it will not be possible to be absolutely sure about the effect of CO_2 .

Experiment XI—Effect of Etching by Water Vapor

Since the attack of water on glass appeared to be the chief cause of fatigue and the chief weakener of glass, some studies of the effect of water upon glass at an elevated temperature were tried. A high pressure autoclave allowing use of steam up to 1000 lb./sq. in. was used to make the tests. It was found that under prolonged treatment at pressures as high as 800 lb./sq. in. all glasses, including silica glass, were attacked severely by water. The soda-lime glasses were attacked the most, Pyrex glasses next, and silica glass least of all. The attack was similar to etching with hydrofluoric acid. In the case of the soda-lime glasses a very thick jelly-like layer formed on the surface. Since the etching was similar to that of hydrofluoric acid, one would expect the rods to get stronger when autoclaved. Unscratched silica rods and rods scratched with coarse emery were autoclaved at 700 to 800 lb./sq. in. for twenty-four hours and then removed and tested for strength (Table XI).

TABLE X.

Test No.	Liquid surrounding rod	Duration of test	Average strength
1	Distilled water	100 sec.	13,400 lb./sq. in.
2	Distilled water saturated with CO_2	100 sec.	13,150 lb./sq. in.
3	Saturated solution of $\text{Ca}(\text{OH})_2$	100 sec.	13,950 lb./sq. in.
4	Normal solution of NaOH	100 sec.	13,750 lb./sq. in.

From these tests it is immediately evident that autoclaving can increase the strength of silica glass, probably by etching out the flaws.

Tests similar to those on the silica rods were made on soda-lime glass rods. In this case the rods were autoclaved for four hours at 600 lb./sq. in. then removed and tested wet in the quick-loading apparatus. The results are shown in Table XII.

These tests on soda-lime rods showed that their strength also could be increased by etching in an autoclave.

Experiment XII

Some further tests were made on the breaking strength of rods in an autoclave. A glass rod was supported by knife edges and loaded centrally with a heavy metal weight. The whole apparatus was put into the autoclave and the steam pressure was applied for about one hour. Then it was observed whether or not the rod had broken. When the rod did not break in the autoclave, it was next tested in the open and found to be very much stronger than comparable rods that had not been autoclaved. But such rods as did break in the autoclave broke at unusually low stresses of 2000 lb./sq. in., and even less, compared with about 5000 lb./sq. in. for the lowest breaking strength of comparable rods when tested for 24 hours in the open. It would appear therefore that glass can fatigue very much more in an autoclave than in ordinary atmospheric conditions.

Discussion of Results of Vacuum Tests

These experiments on the strength of glass in a moderately high vacuum show clearly that the strength of glass is much greater in vacuum than when the glass is wet and in the open, and that fatigue is greatly reduced or entirely eliminated by vacuum treatment.

The increase in strength due to baking in a

TABLE XII.

Test No.	Duration of autoclaving	Pressure of autoclave	Duration of test	Average strength
5	4 hours	600 lb./sq. in.	10 sec.	27,400 lb./sq. in.
6	0		10 sec.	13,000 lb./sq. in.

vacuum is shown by experiments I-IV. The ratio of the vacuum strength to the wet strength is 2.5 for Pyrex, 2.0 for soda-lime glass, 1.7 for silica glass, and 1.5 for porcelain. Thus it would appear that this effect is dependent somewhat on chemical composition. The fact that it is relatively low for porcelain and silica glass (which are chemically resistant) would suggest that the effect is correlated with the chemical attack of moisture or gases at the tip of the crack.

It is quite evident that glass fatigues very little, if at all, in a high vacuum. Tests 2 and 3 of experiment VII show that the strength is 25,800 lb./sq. in. on the 0.1-second test and 25,300 lb./sq. in. on the 10-second test. Also, Tests 1 and 2 of experiment VI give the strength as 25,900 on both the 0.1-second and 10-second tests in high vacuums. The fact that there was very little fatigue at high vacuums was evident from observations with a cathode-ray oscillograph of the times required to break the rods. The rods generally broke as soon as the load came on or not at all; only occasionally was there definite evidence of delayed breakage. However, unless the rods were baked out at 350°C or more, at high vacuum, there was considerable evidence of fatigue. It would therefore appear that fatigue is caused primarily by the action of water vapor and adsorbed gases acting in the flaws. It is very doubtful that even heating the glass up to 350°C or more removed all the water vapor from the flaws, since they were in the neighborhood of 0.005-inch deep in the glass used and probably only a few Angstrom units wide at the tips. Too little theoretical knowledge is available to estimate the length of time required to evacuate the water vapor under those circumstances. It is possible that the most sensitive test of the presence of water vapor in the tip of a flaw is the occurrence of fatigue in strength tests.

There is considerable evidence that the "strength" of glass is a much more complicated

TABLE XI.

Test No.	Time autoclaved	Steam pressure	Duration of test	Average strength
1-Scratched	24 hours	700-800	10 sec.	30,000 lb./sq. in.
2-Scratched	0		10 sec.	6000 lb./sq. in.
3-Unscratched	24 hours	700-800	10 sec.	54,000 lb./sq. in.
4-Unscratched	0		10 sec.	19,000 lb./sq. in.

matter than is generally realized. Such evidence is apparent in a comparison between the tests in experiments VI and VII. In both cases the rods were of the same composition. The rods in experiment VI had been stored for a few years, some of the time in an unheated warehouse, and showed slight weathering. These rods showed a strength of 32,200 lb./sq. in. on the 0.1-second test in a 0.1-mm vacuum, as compared with 25,900 lb./sq. in. for rods baked out in a vacuum of 0.1 micron. Careful checks were made so there was no doubt of the phenomenon, yet it was unrepeatable in experiment VII on glass that was not so old and had been stored in a heated room most of the time.

The only hypothesis to explain these rather erratic results seems to be that gels form in the flaws which strengthen the glass but have much greater fatigue rates than glass. When the glass is thoroughly baked out, the gels crack or lose their strength to an extent that makes them a negligible factor.

The chemical attack of water on glass at elevated temperatures and pressures was shown in experiments XI and XII. It was somewhat surprising to see the rapidity of attack of water on silica glass at steam pressures around 700 to 800 pounds per square inch. That the flaws were etched out and that the strength of the glass was increased was clearly demonstrated. It was surprising to find that the process of fatigue could be remarkably speeded up at these temperatures, so that glass fatigues to much less than its 24-hour strength in a matter of an hour or possibly only a few minutes. These results strengthen our feeling that the process of fatigue is essentially a chemical attack.

The chemical attack upon glass by moisture is extensively discussed in the literature and summarized by Morey.⁵ That a chemical attack inside the flaws should decrease the strength of glass is not surprising, as one would expect the bonds under the greatest tension to be those most ready to rupture and combine to form new compounds. This chemical attack will of course depend on two factors, the first being the ability of the water to penetrate to the tip of the flaw and the second being the chemical durability of

the glass. The fact that all cracks are pretty much alike dimensionally as well as the fact that all glasses, including silica, are attacked by water probably accounts for the similarity of strength between glasses at room temperature. That there are differences in the amount of fatigue in glasses of different compositions and that the ratio between the strength of glass in vacuum and the strength of glass wet is different for different glasses indicate that there is a partial correlation between chemical composition and fatigue of wet glass. It was found² that silica appears to fatigue less than soda-lime and Pyrex, and the Pyrex appears to fatigue even more than soda-lime if one takes into account the fact that the samples of Pyrex glass had an initial compression. Then of course in the vacuum work just cited the ratio of the vacuum strength to the wet strength was greatest for Pyrex, with soda-lime following and, finally, the ratio for silica glass was least of all. In general, Pyrex glass is considered to be more chemically durable than soda-lime, but this is not necessarily true when the attacking medium is basic in nature. In a flaw there is no reason for the alkali to be leached out. Since probably the attack in both Pyrex and soda-lime glass is basic in nature the correlation with chemical durability seems to hold.

These tests on the strength of glass under different conditions of the surrounding medium give considerable insight into the ultimate questions pertaining to the fatigue of glass under constant load. Specifically, it seems to be necessary to distinguish between glass in which the flaws are of recent origin and glass in which the flaws have existed for some time and have been exposed to the atmosphere.

In the first case the glass has a definite strength under high vacuum no matter what the duration of the load (at least to the extent covered in our tests). When even slight amounts of moisture are present the glass loses strength with duration of load, although this loss is slight in moderate vacuum. As the humidity is increased, the loss of strength with duration of load becomes greater, that is, as the absolute vapor pressure of the water is increased a condition arises where fatigue is very rapid. Thus it appears that when the duration of load is decreased the strength of

⁵G. W. Morey, *The Properties of Glass* (Reinhold Publishing Corporation, New York, 1938), Chap. V.

glass approaches its vacuum strength. This suggests that in extrapolating fatigue curves to shorter times it would be unjustified to go to durations which yield strengths greater than the vacuum strength. Moreover, we should expect the fatigue curves, now extended experimentally only to 0.01-second duration, to level off as the duration of load is decreased to around 0.0001 second. In all probability the extrapolation of the present fatigue curves to shorter times is not legitimate.

In the second case, where the flaws have had time to weather and gels have formed inside them, the picture of the short time strength is too complicated to permit any definite statement. The short time strength probably is determined more by the strength of the gels than by the strength of the glass. Some glass may act like freshly scratched glass and some may show behavior characteristics similar to the weathered glass of experiment VI. The very short time strength can be equal to or greater than the vacuum strength, depending on the gels.

In the middle range of durations of load (from 0.01 second to perhaps several years) the decay of strength with time of wet glass at room temperature and atmospheric pressure may be represented by the equation

$$S = [C_1 / \log C_2 t] + C_3.$$

This equation may be regarded as empirical, and other empirical formulae fit the results equally well. One such equation is discussed by Preston and Glathart in the fourth paper of this series, entitled "The Fatigue Modulus of Glass."⁶

Under "atmospheric" conditions, it is likely that there is very little difference in fatigue between glass with weathered flaws and glass with fresh flaws. But this may depend on the way we define "fatigue"; see the paper of Preston and Glathart mentioned above.⁶

If however the conditions of the atmosphere are changed the picture is greatly altered. As the temperature and water vapor pressure are increased the fatigue is rapidly accelerated until at pressures of a few hundred pounds of steam

a decline in strength takes place in a few minutes that might take hours or years at room temperature. However, if the glass is not broken the flaws will etch out in a few hours and the specimen will become very much stronger than before. Ultimately, the attack of the steam will destroy the glass rods completely and so we may consider the strength to be zero for glass subjected to 1000 lb./sq. in. steam pressure for a week or two.

If a glass rod were under constant load and surrounded by water at room temperature, it is exceedingly probable that its strength would behave in a manner similar to that of glass subjected to elevated temperatures in an autoclave. For fairly high stresses the tips of the flaws would tend to round out without much extension and the glass would eventually become stronger. Ultimately, the glass would dissolve and be destroyed by the water, but many thousands of years would be required in the case of chemically resistant glasses.

It is a moot question as to what will happen to glass under load at room temperature in a dry atmosphere (relative humidity between 1 percent and 50 percent). There is evidence of chemical attack on most glasses under these conditions. The glass will fatigue in about the same way as it does when it is wet. Therefore, it is likely that for many glasses the ultimate results will be the same. Under high stresses the flaws will extend until breakage occurs. Under low stresses of about 1000 to 2000 lb./sq. in. the attack of the water will tend to round out the flaws and the glass will slowly get stronger. On the other hand, crystalline material may form, in the case of some of the less durable glasses, tending to force open the flaws and progressively extend them. In such a case the strength of the glass is zero for durations of load as short as a few years.⁴

CONCLUSIONS

In high vacuum, thoroughly baked out glass can be said to have a tensile strength fairly independent of the time of loading, at least for the shorter durations of load. If glass is exposed to moisture it rapidly loses strength under stress. In ten seconds the strength can decay to $\frac{1}{2}$ or $\frac{2}{3}$ of the vacuum strength, and it may decay to $\frac{1}{3}$

⁶ F. W. Preston and J. L. Glathart, *J. App. Phys.* **17**, 189 (1946).

the vacuum strength in 24 hours. This strength decay appears to be caused by a chemical attack in the flaws of the glass by water and possibly by some gases such as CO_2 .

It therefore appears that for very short durations of load the strength of freshly scratched glass approaches the high vacuum strength. For moderate durations of load ranging from 0.01 second to probably several days or so the strength of wet glass or glass under humid conditions approximately obeys the relationship

$$S = [C_1 / \log C_2 t] + C_3,$$

where S is breaking stress, t is time, and C_1 , C_2 , and C_3 are arbitrary constants. For longer durations of load, several years or more, the chemical composition is so important as to be the controlling factor in strength, but there is reason to believe that at low stresses the flaws can etch out and the glass actually become stronger, in the case of chemically resistant glasses, so that glasses that survive for years may survive for several millions of years. Inevitably of course the chemical attack of the atmosphere will disintegrate the glass to a point where the strength

is zero, in the case of silica glass in probably billions of years.

ACKNOWLEDGMENTS

The authors wish to acknowledge the help of Dr. Justin Glathart, of the Preston Laboratories at present (though not at the time of the experiments) for valuable editorial help, both in the matter of checking the ancient records against these present summaries, and in arranging the matter and the diction for the benefit of readers who had no contact with the early work. The authors realize that, in attempting to write at the end of the European war reports that should have been written at its beginning, they are apt to use terminology which the passage of time should have outmoded. Thanks to Dr. Glathart, we believe, however, that serious ambiguity has been avoided, and major errors in transcribing the old results reduced to very few or eliminated altogether. It is hardly to be expected that in the second and third papers no errors, even minor ones, remain, but we believe that the general picture has been presented as accurately as the experiments disclosed it.

The Fatigue Modulus of Glass

J. L. GLATHART AND F. W. PRESTON
Preston Laboratories, Butler, Pennsylvania
(Received November 1, 1945)

The decline in the breaking strength of glass with an increase in the length of time it is under stress has been noted by various experimenters. Based on the experimental results of T. C. Baker, an empirical relationship is obtained which indicates that the reciprocal of the breaking stress is a linear function of the logarithm of the duration of the stress. Certain implications of this relationship are discussed.

I. INTRODUCTION

VARIOUS experimenters have shown that a glass specimen may sustain a certain static stress momentarily but will ultimately fail under the same stress, with the mere passage of time. No complete theoretical basis for this effect has as yet been established, although Murgatroyd¹ has recently proposed a "viscous pocket" theory for the constitution of glass which, he states, is capable of accounting qualitatively for the observed decrease in strength with time.

The object of the present paper is to present an empirical relation which fits the strength-time curves of all the glasses for which adequate experimental data are available. Some of the implications of this relation are discussed.

II. THE STRENGTH-TIME RELATION

In another paper² T. C. Baker has given his experimental results of the fatigue effect in fused silica, three kinds of commercial glasses, and three porcelains. His results are summarized in Table I.

The total range of durations of stress is seen to be from 0.01 second to 24 hours (86,400 seconds), a range of nearly ten million to one. This is a tremendous range, and it becomes virtually necessary to plot it on a logarithmic base. The various ranges of strengths, however, are relatively small, from 23,880 to 11,740 lb./sq. in. in the case of fused silica, for instance, and from 19,900 to 6,450 lb./sq. in. in the case of soda-lime glass. It is natural to plot this on a linear or arithmetical scale. This method of plotting, f against $\log t$, was the one adopted by Smekal³

who, over a smaller range of durations, reported a linear relationship. This came about, as James Bailey pointed out later,⁴ because Smekal inadvertently misplaced some of his points by a factor of ten. Had they been correctly plotted, his line would not have been straight.

Baker adopted the same rather natural method of plotting, and obtained very definitely curved lines for all of his materials, the curvature being the more obvious because of his longer range of time intervals. Baker, with commendable caution, refused to indulge in any speculation regarding the strengths of his materials beyond the experimental time range, but warned that any point on any of his curves might be in error by as much as ± 10 percent.

The Preston Laboratories have received various inquiries and suggestions regarding this matter of the fatigue of glass, especially since the

TABLE I. Average breaking stress in pounds per square inch (and its reciprocal)* for various durations of load.

Material	0.01 sec.	0.1 sec.	1.0 sec.	10 sec.	100 sec.	1000 sec.	86,400 sec.
Disannealed Pyrex (dry)	24,800 (4.06)	19,800 (5.10)	17,100 (5.85)	15,900 (6.30)	—	12,800 (7.81)	11,800 (8.55)
Scratched Pyrex (wet)	12,900 (7.75)	10,600 (9.43)	9,000 (11.11)	8,250 (12.12)	—	5,750 (17.4)	5,050 (19.8)
Annealed soda-lime glass (wet)	19,900 (5.03)	15,900 (6.29)	—	10,500 (9.53)	—	7,950 (12.58)	6,450 (15.50)
Annealed lead glass (wet)	—	19,000 (5.26)	14,900 (6.71)	11,200 (8.93)	—	9,400 (10.63)	7,950 (12.58)
Fused silica (wet)	—	23,880 (4.19)	—	19,800 (5.05)	15,660 (6.39)	14,280 (7.00)	11,740 (8.53)
Porcelain A (dry)	15,800 (6.34)	13,900 (7.20)	12,300 (8.13)	—	9,700 (10.32)	—	9,100 (11.00)
Porcelain B (dry)	15,400 (6.50)	13,500 (7.41)	11,900 (8.40)	9,900 (10.10)	9,400 (10.64)	9,100 (11.00)	8,900 (11.23)
Porcelain C (dry)	10,900 (9.17)	9,400 (10.64)	8,400 (11.91)	8,000 (12.50)	—	6,000 (16.67)	5,900 (16.95)

* Reciprocals are given in parentheses, and are to be multiplied by 10^{-5} .

⁴ J. Bailey, private communication.

¹ J. B. Murgatroyd, *J. Soc. Glass Tech.* **28**, 406 (1944).

² T. C. Baker and F. W. Preston, "Fatigue of glass under static loads," *J. App. Phys.* **17**, 170 (1946).

³ A. Smekal, *Ergeb. d. exakt. Naturwiss.* **15**, 106 (1936).

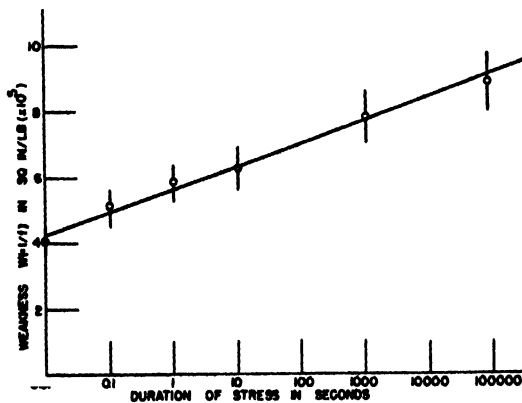


FIG. 1. Disannealed Pyrex glass, dry.

publication of the Philadelphia lecture,⁵ and the matter has recently been considered by Murgatroyd and Haward.⁶ In view of this interest, it seems desirable to discuss Baker's curves a little more fully with respect to whatever empirical information may be obtained from them.

We have known for some time that if $1/f$, the reciprocal of the breaking stress, is plotted against $\log t$ a straight line can be drawn which, for every one of Baker's materials, passes within the "possible error" he assigned to his values. A note to this effect appeared in *Nature*.⁷ This relation has also been noted by Haward,⁶ whose observations reached us after the note in *Nature* had been published.

Figures 1 to 8 show the straight lines that result when Baker's values are plotted in this way. Figure 9 is a similar plot of some previous data reported⁸ from this laboratory, which were obtained on plate glass laths of dimensions $16'' \times 2'' \times \frac{1}{4}''$, considerably larger than Baker's round rod specimens. Although they cover a much shorter time range, these data are included because they were cited by Haward.⁶

In the case of two materials, fused silica and soda-lime glass, this reciprocal relationship is so exact that nearly every point lies on the straight line. In the case of the Pyrex glass and the lead glass the fit is also quite good, but the points for the porcelains deviate considerably, although even in these cases straight lines can be drawn within the errors assigned. The discussion pre-

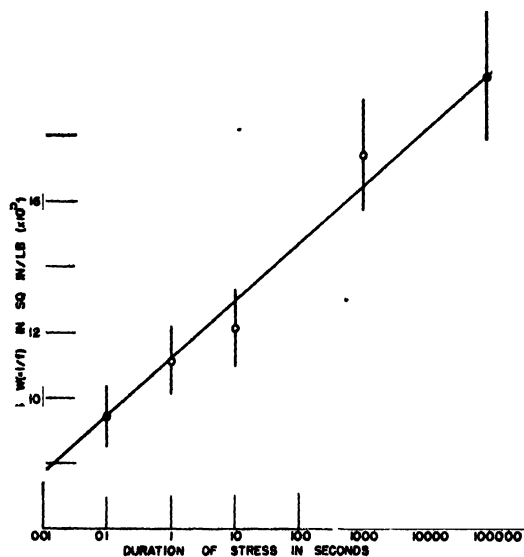


FIG. 2. Disannealed Pyrex glass, scratched, wet.

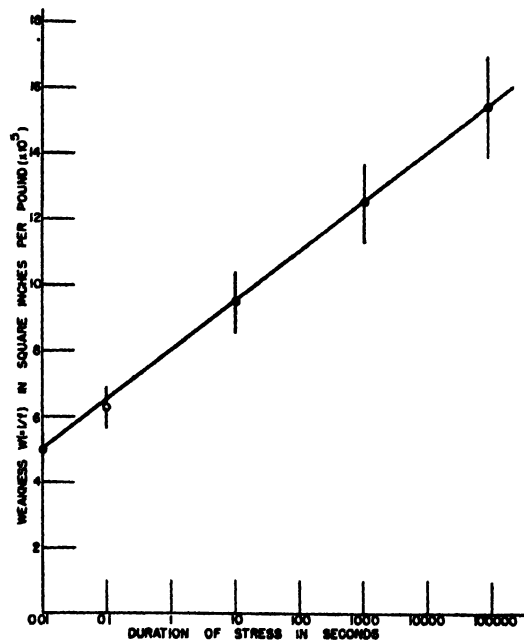


FIG. 3. Annealed soda-lime glass, wet.

sented in this paper will apply more particularly to the four glasses, Pyrex, soda-lime, lead, and silica.

Since the straight lines fit so well in these cases, it is natural to inquire as to their equations, and whether any theoretical basis can be found for them. We find that if there is any theoretical basis at all it is expressed in terms of concepts such as "weakness," rather than strength, and a "fatigue modulus."

⁵ F. W. Preston, *J. App. Phys.* **13**, 623 (1942).

⁶ R. N. Haward, Comment on Murgatroyd's paper. See reference 1, p. 427.

⁷ F. W. Preston, *Nature* **156**, 55 (July 14, 1945).

⁸ F. W. Preston, *J. Am. Cer. Soc.* **18**, 220 (1935).

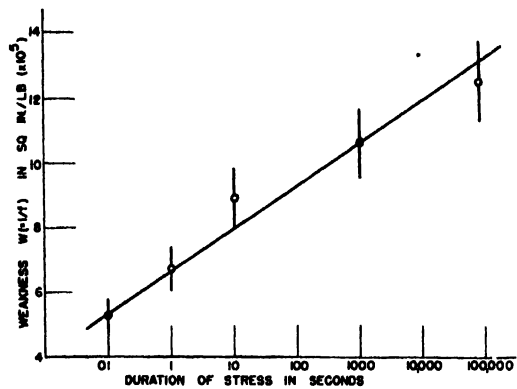


FIG. 4 Annealed lead glass, wet.

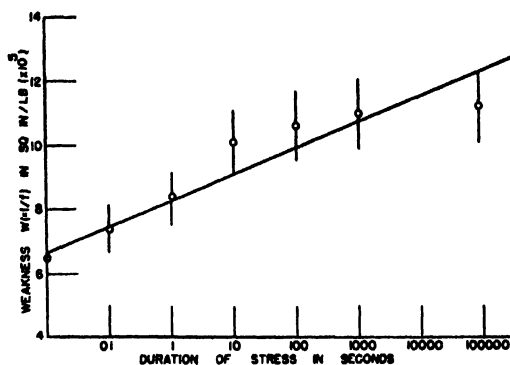


FIG. 7. Porcelain, type B, dry.

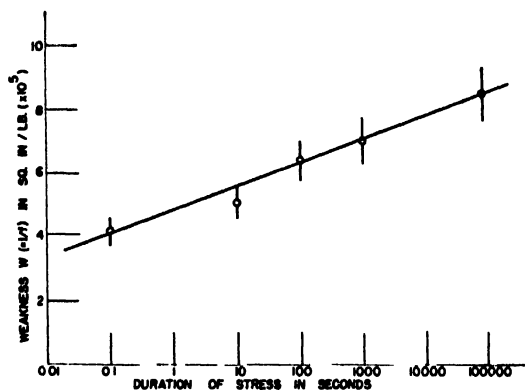


FIG. 5. Fused silica, wet.

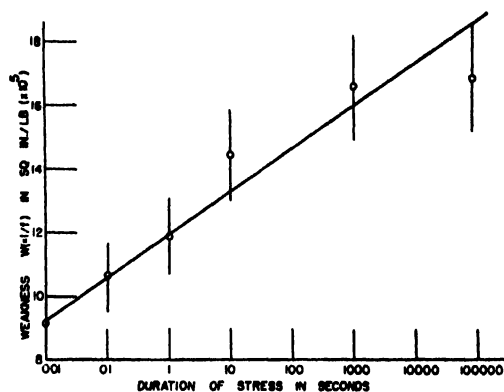


FIG. 8. Porcelain, type C, dry.

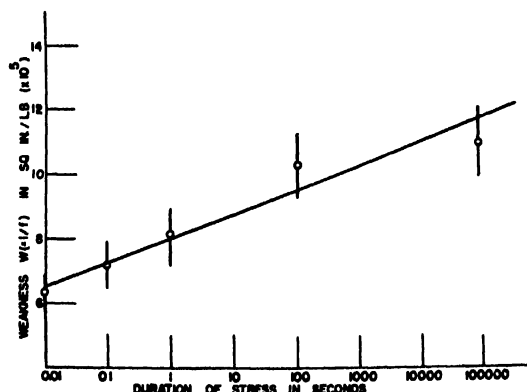


FIG. 6. Porcelain, type A, dry.

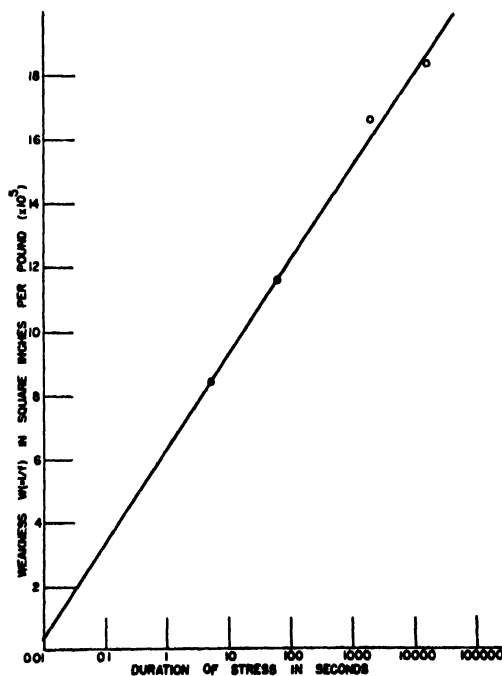


FIG. 9. Plate glass, dry.

It is more or less inevitable, also, that we should speculate on the extrapolation of the curves, to time intervals both longer and shorter than were attained in Baker's experiments. There is a practical necessity for speculating on the longer times, and fortunately there is no apparent reason why the curves should not continue as straight lines to times far beyond anything of

technological interest. At the short time end things are more complicated.

III. EXTRAPOLATIONS

The various curves (straight lines) which represent Baker's results are expressed by equations of the form

$$1/f = a + m \log t, \quad (1)$$

where f is the breaking stress, t is the time, and a and m are constants. Each line is thus defined by its slope $d(1/f)/d(\log t)$ and by the point (t_0) , where the line intersects the time axis.

With such materials, once the fact that they plot to straight lines is conceded, the question arising in the mind of the experimenter—or reader—is apt to be: "how far can the curves be extrapolated at each end with any confidence?"

(a) The Long Time End

The equation above would make the strength zero at infinite time. This is psychologically satisfying, and, furthermore, is required by some theories of strength which have been recently developed. Otherwise stated, it means that even a slight stress will break the glass if only it lasts for a sufficiently long time.

There seems to be nothing in the physics of long time loading that would cause us to deny the possibility that the straight-line law holds more or less indefinitely, provided temperature and humidity are constant and there is no physical change in the specimen other than changes which operate during our relatively brief loading intervals. The straight-line relation holds from 0.01 second to nearly 10^5 seconds, a range of time durations of 10^7 to 1. Our experiments end at a duration of one day, but there seems no reason why we should expect any departure from the linear relation if we extend the time range by a factor of 10^3 , that is, from one day to three years, or even by a factor of 10^4 , that is, up to thirty years. This is much longer than most bottles last, and for most of their lives they are not under stress. There are not many examples of plate glass, or other kinds of glass, that are expected to last, under stress, for so great a length of time.

If the curve were extended to the right, beyond one day, as far as we have determined it experi-

mentally to the left, that is, in the ratio of 10^7 to 1, we arrive at a period of 30,000 years, or roughly the time that has elapsed since the maximum of the last ("Wisconsin") glaciation, and several times as long as all recorded history.

In Table II below we take the strength at the end of one day of stress as unity and calculate, for fused silica and soda-lime glass, the proportionate strengths at several other longer durations, on the assumption that this linear relationship holds indefinitely as we move to the right.

(b) The Short Time End

The strength-time equation would make the strength infinite for some very brief, but not infinitesimally short, duration (t_0). This point is obtained by extending the straight line of the graph to the left until it intersects the time axis. The values of t_0 for the various materials tested are shown in Table III.

This threshold time, t_0 , is obtained from the equation of the curve and its value depends, therefore, on the way the straight line is drawn among the experimental points. It is interesting to note, however, that for all those glasses which were tested wet, the value of t_0 is in the neighborhood of 1 microsecond, whereas its value for those materials which were tested dry is very much less. However, other factors than moisture are involved, of course. For example, porcelains *A* and *B*, which had identical compositions but different firing temperatures, have nearly the same value of t_0 , whereas porcelain *C*, of different composition, has a t_0 value 20 to 40 times as large.

TABLE II. Extrapolation at the long time end

Material \ Duration	1 day	1 year	10^3 years	10^4 years	2×10^5 years	Infinite time
Fused silica	1.0	0.82	0.68	0.57	0.50	0
Soda-lime glass	1.0	0.80	0.65	0.55	0.46	0

TABLE III. The threshold time.

Material	t_0 (sec.)	Material	t_0 (sec.)
Soda-lime glass	5.0×10^{-6}	Pyrex (dry)	0.0085×10^{-6}
Lead glass	10.2×10^{-6}	Porcelain <i>A</i>	0.00004×10^{-6}
Fused silica	0.2×10^{-6}	Porcelain <i>B</i>	0.00008×10^{-6}
Pyrex (wet)	0.4×10^{-6}	Porcelain <i>C</i>	0.0016×10^{-6}

TABLE IV. Theoretical strengths at short durations.

Material \ Duration	10 ⁻² sec.	10 ⁻³ sec.	10 ⁻⁴ sec.	10 ⁻⁵ sec.	10 ⁻⁶ sec.
Fused silica	28,200	37,000	50,700	80,700	196,000
Soda-lime glass	19,900	28,800	51,300	232,000	∞

At present, t_0 is merely a mathematical abstraction. It could not be even remotely approached experimentally with such a set-up as we were using. Yet on our logarithmic scale it is not very far beyond the short time end of the experimentally determined curve. In fact, for some of the materials it is much closer to the left-hand end of the experimental curve than this left-hand end is to the right-hand end, so the extrapolation, considered as a mathematical operation, is not too extravagant.

Physically, however, Baker probably carried the curve toward the short time end as far as it can be carried unless some totally different and much smaller set-up can be devised. Possibly some arrangement using a silver plated glass fiber supported on a short span in a magnetic field, and deflected by the discharge of a condenser through the silver plating, might serve to reach shorter time intervals.

Let us assume that by some means or other the curve could be extended, experimentally, to shorter times, and that it did, in fact, continue straight for some distance. The calculated strength values of fused silica and soda-lime glass at several shorter times are shown in Table IV.

The behavior of glass under very high stresses cannot be expected to parallel its behavior under moderate stresses, because we cannot expect Hooke's law to be obeyed at extreme strains. It seems not inconceivable that the straight-line law might apply fairly well until stresses of a million pounds per square inch are reached, with a resulting strain or extension of 10 percent in the length of the fiber. But it is scarcely likely that it could apply at stresses of ten million pounds per square inch, a figure equal to Young's modulus as calculated for small strains.

At the short time end the inertia of the material becomes an important factor and makes experimentation impractical, by any method as yet envisaged, much beyond the point to which Baker carried it.

Thus, for more reasons than one, the shape, or

even the physical reality, of the curves beyond Baker's short time values must remain doubtful for the present, although the theoretical implications may make us wish to find means of extending them in that direction. This might be accomplished by totally different techniques, perhaps by impact testing, as hinted by Baker.

IV. THE CONCEPT OF "STRENGTH" OF GLASS

From the empirical equation

$$f e t_0 / f = \log_e (t / t_0), \quad (2)$$

which can be deduced from Baker's experimental results, it can be seen that the strength f is not a constant of the material but that it depends upon a variable (time, t) and on two constants of the material (or at least of the specimen). One of these is t_0 , which is the maximum duration of stress for which the strength is theoretically infinite, and the other is $f e t_0$, which is the strength of the glass for a duration of stress $e t_0$, where $e = 2.718$, the base of natural logarithms.

We may tentatively call t_0 the "threshold duration." It is the least interval of time in which a tensile stress, however large, can rupture the material (or the specimen). In our experiments on glass it is of the order of a microsecond, and in this time a crack can be propagated no more than $60,000 \times 10^{-6}$ inches = $\frac{1}{8}$ inch, and the stress pulse no more than $\frac{1}{8}$ inch. These dimensions seem to bear no relation to the ultimate structure of the material, but they are comparable with the shortest dimension of the specimen (0.22 inch), or with the shortest dimensions of an effective crack. They may even correspond somewhat closely with the radius of the smallest "mirror" (initiation crack) that can be produced under infinite stress. This in turn may be, and probably is, a function of the size of the specimen, and not a true constant of the material. Obviously, the mirror on a fine fiber has to be smaller than $\frac{1}{8}$ inch.

If t_0 is a property of the specimen rather than the material, then so is $e t_0$, and likewise, presumably, is the strength at this time interval, or $f e t_0$.

Equation (2) above merely expresses an equality of ratios; it describes the "fatigue," but it says nothing about strength itself. The strength at one duration is expressed only in terms of the

strength at some other duration, not in terms of the physical constants of the material, or the dimensions of the specimen. Until t_0 and $f_{e t_0}$ can be so described, the equation is limited in exactly the same way as the experiments upon which it is based.

By differentiation of (2) we get

$$f_{e t_0} \frac{d}{dt} \left(\frac{1}{f} \right) = \frac{1}{t}. \quad (3)$$

This eliminates t_0 from the picture but does not simplify things very much. The reciprocal of the breaking stress, that is, $1/f$, may be called the "weakness" of the specimen; it is the area that must be assigned to carry unit load, if the specimen is to last out the time interval t . If this is denoted by w , then

$$\frac{1}{f} \equiv w \quad \text{and} \quad \frac{1}{f_{e t_0}} \equiv w_{e t_0}. \quad (4)$$

So

$$dw/dt = w_{e t_0} (1/t). \quad (5)$$

The rate of increase of weakness, dw/dt , we may call "fatigue." If so, Eq. (5) shows that fatigue is a property which varies inversely with the duration of stress.

Equation (2) may now be written

$$t/t_0 = e^{w/u_{e t_0}}, \quad (6)$$

and we then have, from Eq. (5),

$$\frac{dt}{dw} = \frac{t_0}{w_{e t_0}} e^{u_{e t_0} w}. \quad (7)$$

t is the "lastingness," or endurance, of the specimen and is here expressed in terms of w , the area assigned to carry unit load.

We may write, finally

$$\frac{1}{t} \frac{dt}{dw} = \frac{1}{w_{e t_0}} = f_{e t_0}. \quad (8)$$

This says that the relative gain in endurance (as compared with the previous total endurance), due to unit increase in assigned area, is a constant of the material, or at least of the specimen.

In other words, if we want to increase the endurance by 1 percent we must add a constant number of square inches of cross-sectional area to

TABLE V. Fatigue modulus.

Material	Modulus in. ² /lb.	Material	Modulus in. ² /lb.
Pyrex (dry)	3.0×10^{-6}	Fused silica	3.2×10^{-6}
Pyrex (wet)	1.6×10^{-6}	Porcelain A	3.3×10^{-6}
Soda-lime glass	6.5×10^{-6}	Porcelain B	2.4×10^{-6}
Lead glass	5.7×10^{-6}	Porcelain C	5.9×10^{-6}

carry unit load. The area, dw , to be added is inversely proportional to the "discriminant strength," $f_{e t_0}$, since, from Eq. (8), $f_{e t_0} \times dw$ is constant for a given value of dt/t .

If, instead of talking about a 1 percent increment in the life of the specimen, we define "fatigue" as that number of square inches that must be added to the cross-sectional area in order to double the life of the specimen, then $\delta t = t$ in Eq. (8) and

$$\delta w = w_{e t_0} = 1/f_{e t_0}. \quad (9)$$

In this case, δw is the fatigue in square inches per pound of load.

For the various glasses tested the δw values in in.²/lb. are given in Table V.

It should be understood that δw is to be deduced from Eq. (8), not ascertained by determining experimentally what area must be added to a specimen to increase its life from, say, one minute to two minutes. δw is a kind of "modulus," in the sense that Young's modulus E is "that stress which would double the length of the specimen." In practice, we do not try to double the length of the specimen. Rather, we use a small increment of length, assume Hooke's law, and deduce E from the equation

$$(1/L)(dL/df) = 1/E.$$

This compares with our Eq. (8)

$$(1/t)(dt/dw) = 1/w_{e t_0}.$$

Subject to this understanding, it seems we can say: Fatigue is that number of square inches that must be added to the cross-sectional area assigned to carry unit load (one pound) in order to double the life of the specimen.

V. PHYSICAL CONSIDERATIONS

The above discussion consists merely in the manipulation of the mathematical symbols in an attempt to interpret the straight-line relationship

derived from Baker's curves in terms of concepts which may find further application in this field.

There are other considerations however, some of which have been outlined. Another is the inherent unlikeliness that the lines representing the soda-lime glass and the silica glass actually cross one another to the left of the observed points, which would mean that the soda-lime glass becomes stronger than fused silica for stresses of sufficiently short duration. It seems more likely that, if the curves have any physical reality when extrapolated far to the left, they do not remain straight, but presumably flatten rapidly and become asymptotic to the time axis.

In all probability, this means in turn that our experimental straight-line relationship "in the visible" is itself only an approximation, even though it seems an extraordinarily good one in at least some of the instances.

Haward's theory⁶ leads to an equation

$$S(e^{P_0/t} - e) = kt,$$

where S is the applied stress, t is the time required for the glass to break under this stress, and P_0 and k are constants. For $t=0$, this yields $S=P_0$, that is, the strength is not infinite, as would be required if our straight line held indefinitely.

On the other hand, Murgatroyd's viscous

pocket theory¹ implies a relation identical with our Eq. (1). It is obvious that much more theoretical work needs to be done on the matter of the meaning of the "strength" of glass and on the relation of strength to time.

CONCLUSIONS

The ordinary concepts of "strength" do not apply to glass, at any rate not to glass as we know it, exposed to the earth's atmosphere and to atmospheric moisture. Under such normal conditions the "life" or "endurance" of a stressed glass specimen depends on the cross-sectional area which we assign to carry unit load. The greater the area we assign, the longer the life of the specimen, the life, or endurance, varying exponentially with the assigned area.

This permits the calculation of a certain constant, which we have called the "fatigue modulus," which describes the behavior of the specimen under the particular conditions of the experiment.

However, in all probability, this modulus is not so much a property of the substance of the specimen as it is a function of the surrounding atmosphere or, more accurately, of the conditions obtaining at the surface of the specimen.

Induction Heating of Hollow Metallic Cylinders

ANDREW GEMANT

The Detroit Edison Company, Detroit, Michigan

(Received December 14, 1945)

This paper deals with the heat input to hollow cylindrical metal cores subjected to induction heating. The rigorous expression for the heat input, valid for the whole frequency range, is developed and is presented by Eq. (27) of this paper. It is shown in an example how the equation is used for numerical computations. Existing approximate formulas for the low frequency and the high frequency range are checked by means of the rigorous equation; the maximum deviation between the rigorous and approximate equations is about 10 percent.

1. THE PROBLEM

AS a result of the growing number of practical applications of induction heating of metals, many papers have been published that present the theory of induction heating and its usefulness in solving practical problems. One of the chief applications of the theory involves the calculation of the heat input to a metal core surrounded by a coil carrying alternating current. In the

case of a solid core rigorous equations are usually given, from which are deduced the simplified equations that hold with good approximation, if the frequency is either higher or lower than certain critical limits.

When hollow cores are discussed, the rigorous equations referring to that case are not given. It is occasionally pointed out how these may be obtained, but the final equation is not presented.

Instead, only the approximate equations for frequencies above and below certain limits are given. This type of treatment can be found, for instance, in papers by Greig,¹ Wall,² and Baker,³ as well as in a book by Warren.⁴

While the approximate equations just referred to seem to be sufficient for practical purposes, it appeared to the author that a presentation of the rigorous equations would be valuable for the following reasons:

(1) The practical importance of the problem certainly justifies knowledge of the rigorous solution.

(2) A reliable check on the validity of approximate solutions can best be made by means of the rigorous solution.

(3) It might well happen that certain variants of the hollow-core problem will be found for which the approximate solution is not applicable, in which case the knowledge of the rigorous solution would prove valuable.

The complete solution for the heat input to hollow cores in induction heating processes is presented in the following. First, a development of the equation is given, and next a numerical application is presented.

2. SOLUTION OF THE DIFFERENTIAL EQUATION

For the following analysis, the unit length of a hollow cylindrical metal core that is located coaxially within a coil is considered. An annulus of the metal of radius r and thickness dr is now considered. If the magnetic field parallel to the axis of the core is H (instantaneous value), then the total flux ϕ within the radius r is

$$\phi = 2\pi \int_0^r \mu H r dr, \quad (1)$$

where μ , the magnetic permeability, drops at the inner radius of the core from its value for the metal to unity. The electromotive force E along the circumference of the annulus is (with t = time)

$$E = -\partial\phi/\partial t. \quad (2)$$

The current density G along the circumference of

the annulus is

$$G = E/2\pi\rho r, \quad (3)$$

where ρ = electrical resistivity. Finally, the gradient of the magnetic field intensity H is

$$\partial H/\partial r = -4\pi G. \quad (4)$$

These four equations contain four variables: H , ϕ , E , and G as functions of r and t . Elimination of three of the four permits the setting up of the differential equation for any one of the four. For H , one has:

$$\frac{\partial^2 H}{\partial r^2} + \frac{1}{r} \frac{\partial H}{\partial r} = \frac{4\pi\mu}{\rho} \frac{\partial H}{\partial t}. \quad (5)$$

Since the alternating current in the coil is sinusoidal, H can be replaced by $He^{2\pi i f t}$, where H now stands for the complex amplitude and f = frequency of the alternating current. Similarly, $\partial H/\partial t$ can be replaced by $2\pi i f H e^{2\pi i f t}$. Equation (5) then takes the form

$$\frac{d^2 H}{dr^2} + \frac{1}{r} \frac{dH}{dr} - ik^2 H = 0, \quad (6)$$

where

$$k^2 = 8\pi^2 \mu f / \rho. \quad (7)$$

In the following H stands for r.m.s.-value instead of amplitude; the form of Eq. (6) is obviously not affected by this change. The solution of Eq. (6) can be found from the solution⁵ of the differential equation of the general form:

$$\frac{d^2 y}{dx^2} + \frac{1}{x} \frac{dy}{dx} + m e^{nx} y = 0$$

(m and n = constants), whose solution is:

$$y = Z_0(m^{1/2} e^{n/2x}),$$

where Z_0 is the sum of two Bessel functions of the first and third kind, both of zero order and each multiplied by an arbitrary complex constant. Substituting $-k^2$ for m and $\pi/2$ for n , the solution of Eq. (6) becomes

$$H = Z_0(i^{1/2} k i r), \quad (8)$$

in which Z_0 , as stated before, is the sum of J_0 and $H_0^{(3)}$, Bessel functions of the first and third kind, each multiplied by a constant. It is convenient to introduce the so-called modified Bessel

¹ Greig, *World Power* 25, 246 (1936).

² F. F. Wall, *Engineer* 173, 209 (1942).

³ R. M. Baker, *Trans. A.I.E.E.* 63, 273 (1944).

⁴ A. G. Warren, *Mathematics Applied to Electrical Engineering* (Chapman & Hall, London, 1941).

⁵ E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943).

functions according to the definitions

$$I_0(x) = J_0(ix)$$

and

$$K_0(x) = -\frac{\pi i}{2} I_0^{(1)}(ix).$$

so that Eq. (8) becomes

$$H = A I_0(ikr) + B K_0(ikr), \quad (9)$$

where A and B are the complex constants of integration.

By using Eq. (9), the other essential variables, namely the current density G and the electromotive force E can be computed. Since both are sinusoidal and have the same frequency f as H , the form of Eqs. (3) and (4) is not changed if G and E from now on stand for the complex r.m.s. values, instead of instantaneous ones. From Eq. (4):

$$G = -\frac{ikA}{4\pi} I_0'(ikr) - \frac{ikB}{4\pi} K_0'(ikr), \quad (10)$$

where the prime indicates differentiation after the argument. Further, from Eq. (3):

$$E = -\frac{ikr\rho A}{2} I_0'(ikr) - \frac{ikr\rho B}{2} K_0'(ikr). \quad (11)$$

Eqs. (9)–(11) give the essential variables as functions of the radius r within the metal part of the hollow core.

3. COMPUTATION OF THE CONSTANTS OF INTEGRATION

The next step is the evaluation of the constants of integration A and B from the boundary conditions. The two boundary conditions are:

(1) For $r=a$ (a =radius of the outer surface of the hollow cylinder), H assumes the value H_0 which is the field inside the coil that is unaffected by the metal core.

(2) For radii from zero to b (b =inner radius of the hollow cylinder) the field is constant and equal to H_0 , the value that obtains from Eq. (9) for $r=b$. The first condition yields the equation:

$$H_0 = A I_0(ika) + B K_0(ika). \quad (12)$$

The second condition can be expressed in the following manner. First, H_b is written:

$$H_b = A I_0(ikb) + B K_0(ikb). \quad (13)$$

The total flux within the air space, ϕ_b , is then $b^2\pi H_b$. Further: $\partial\phi_b/\partial t = i2\pi^2fb^2H_b$. But this must be equal to $-E_b$, where E_b , the e.m.f. at $r=b$, is given by Eq. (11), if r is replaced by b . Hence, the second boundary condition yields the equation:

$$A \left[\frac{k\rho}{2} I_0'(ikb) - (\sqrt{i}) 2\pi^2fb I_0(ikb) \right] + B \left[\frac{k\rho}{2} K_0'(ikb) - (\sqrt{i}) 2\pi^2fb K_0(ikb) \right] = 0. \quad (14)$$

The constants A and B can be evaluated in terms of H_0 , a , and b from Eqs. (12) and (14). At this stage of the calculation, it is convenient to replace the Bessel functions by their real and imaginary components. The following notations are customary:

$$I(i^{\frac{1}{2}}x) = ber\ x + i\ bei\ x$$

and

$$K_0(i^{\frac{1}{2}}x) = ker\ x + i\ kei\ x.$$

The corresponding differential forms are:

$$I_0'(i^{\frac{1}{2}}x) = i^{-\frac{1}{2}} ber' x + i^{\frac{1}{2}} bei' x$$

and

$$K_0'(i^{\frac{1}{2}}x) = i^{-\frac{1}{2}} ker' x + i^{\frac{1}{2}} kei' x.$$

In order to obtain A and B in convenient forms, the following notations are introduced:

$$\begin{aligned} u &= ber\ ka, \\ v &= bei\ ka, \\ w &= ker\ ka, \\ s &= kei\ ka, \\ p &= 2\ bei' kb - kb\ ber\ kb, \\ q &= -2\ ber' kb - kb\ bei\ kb, \\ s &= 2\ kei' kb - kb\ ker\ kb, \\ t &= -2\ ker' kb - kb\ kei\ kb. \end{aligned} \quad (15)$$

Evaluation of the constants A and B can now be carried out. The result is:

$$A = H_0 \frac{(s+it)}{(u+iv)(s+it) - (p+iq)(w+is)} \quad (16)$$

and

$$B = H_0 \frac{-(p+iq)}{(u+iv)(s+it) - (p+iq)(w+is)}. \quad (17)$$

Both complex constants can be written as:

$$A = A_r + iA_i, \quad (18)$$

and

$$B = B_r + iB_i, \quad (19)$$

where the four real constants are given by Eqs. (16) and (17); numerical computation of these four constants is carried out according to standard methods.

It is easy to verify that for $b=0$, the solid core, B , becomes zero. The result for H in Eq. (9) is then given by the I_0 function alone.¹

It may be added that the present treatment holds not only for the case of a hollow metallic cylinder, but also for the general case of a solid core surrounded by a solid cylinder of another material. ϕ_b is then not simply $b^2\pi H_b$, but a more involved function of b and the constants of the solid core, as known from the equations valid for the solid cylinder. Since ϕ_b is still linear in H_b , Eq. (14) will have the same general form, and only the second terms in the expressions (15) will be correspondingly more involved.

4. CALCULATION OF THE HEAT INPUT

The heat input into the hollow cylinder is calculated from the current density—Eq. (10)—in conjunction with Eqs. (18) and (19). The heat input per unit length W is given by its value $|G|^2\rho$ for unit volume, integrated over the cross-sectional area of the cylinder:

$$W = 2\pi\rho \int_0^a |G|^2 r dr. \quad (20)$$

In order to solve this integral, it may be conveniently written in the form

$$W = 2\pi\rho \int_0^a G\bar{G}r dr, \quad (21)$$

in which $\bar{G}$ is the conjugate complex value of G . This integral can be solved by making use of the following general integral relation in which Y_p and Z_p denote two different Bessel functions of p th order, both the sum of two functions of the first and third kind:

$$\int Y_p(ax)Z_p(\beta x)x dx = \frac{\beta x Y_p(ax)Z_{p-1}(\beta x) - \alpha x Y_{p-1}(ax)Z_p(\beta x)}{\alpha^2 - \beta^2}.$$

Here α and β are constants.

$Y_p(ax)$ is now identified with G , $Z_p(\beta x)$ with $\bar{G}$, and x with r of Eq. (21). The four functions appearing on the right-hand side of the above equation, as well as α and β , can then be easily evaluated, as follows.

Concerning $Y_p(ax)$, it is given by Eq. (10), or, recalling the general relation

$$Z_0'(x) = -Z_1(x)$$

by

$$Y_p(ax) = \frac{ikA}{4\pi}I_1(ikr) + \frac{ikB}{4\pi}K_1(ikr). \quad (10a)$$

From this it follows that $p=1$ and $\alpha=ik$.

Expressing $Y_p(ax)$ in the form of tabulated functions, Eq. (10) is converted, by using the form of I_0' as given in Section 3, into:

$$Y_p(ax) = -\frac{kA}{4\pi}(ber' kr + i bei' kr) - \frac{kB}{4\pi}(ker' kr + i kei' kr). \quad (22)$$

Concerning $Z_p(\beta x)$, or $\bar{G}$, it should be recalled that the conjugate value of any function is given by replacing all i 's occurring in it by $-i$. Hence, from (10a) it follows that $\beta=(-i)k$ and from (22):

$$Z_p(\beta x) = -\frac{k\bar{A}}{4\pi}(ber' kr - i bei' kr) - \frac{k\bar{B}}{4\pi}(ker' kr - i kei' kr), \quad (23)$$

where $\bar{A}$ and $\bar{B}$ are given by Eqs. (18) and (19) with $-i$ instead of i .

As to $Y_{p-1}(ax)$, Eq. (10a) is used in replacing I_1 by I_0 and K_1 by K_0 . It then follows that

$$Y_{p-1}(ax) = \frac{kA}{4\pi}(i^{\frac{1}{2}} ber kr - (-i)^{\frac{1}{2}} bei kr) + \frac{kB}{4\pi}(i^{\frac{1}{2}} ker kr - (-i)^{\frac{1}{2}} kei kr) \quad (24)$$

and

$$Z_{p-1}(\beta x) = \frac{k\bar{A}}{4\pi}[(-i)^{\frac{1}{2}} ber kr - i^{\frac{1}{2}} bei kr] + \frac{k\bar{B}}{4\pi}[(-i)^{\frac{1}{2}} ker kr - i^{\frac{1}{2}} kei kr]. \quad (25)$$

Equations (22) to (25) represent the four functions to be substituted into the numerator of the right-hand side of the general integral given above. The denominator is given by

$$\alpha^2 - \beta^2 = 2ik^2.$$

In carrying out the substitution indicated, the four constants A_r , A_i , B_r , and B_i are found to

$$W = \frac{\rho}{8\pi b} \left| \begin{aligned} &kr\kappa(ber\ ber' + bei\ bei') + kr\lambda(ker\ ker' + kei\ kei') + kr\mu(ber\ ker' + bei\ kei') \\ &+ ber' ker + bei' kei) + kr\nu(bei\ ker' - ber\ kei' + bei' ker - ber' kei) \end{aligned} \right| \quad (26)$$

where, for sake of simplicity, the argument kr after each function is omitted.

It can be seen that, for the solid core for which $B=0$, only the first of the four terms within the bracket remains, giving the known solution for this case.

5. NUMERICAL CALCULATION OF A SPECIAL CASE OF HOLLOW CORE

In order to show the use of Eq. (27), a numerical example treated by R. M. Baker³ will be discussed. That author calculated the heat input by means of two approximate equations and a check of these, as explained in Section 1 of the present paper, can now be carried out by using the rigorous solution as given. It will be shown in the course of this section that the approximate equations are in good agreement with the rigorous one.

For low frequencies, characterized by the condition that the depth of current penetration is larger than the wall thickness, Baker uses the formula

$$W = \frac{61.8 \times 10^{-16} f^2 a^2 t_w \rho H_0^2}{\rho^2 + 15.5 \times 10^{-16} f^2 a^2 t_w^2},$$

where W , the heat input per unit length, is given in watts/cm and t_w , the wall thickness, is measured in cm, ρ is measured in ohms cm, and $\mu=1$ (non-magnetic metal).

For high frequencies, characterized by the condition that the depth of current penetration is smaller than the wall thickness, the formula

$$W = 7.92 \times 10^{-16} a(f\rho)^3 H_0^2$$

is used. At the frequency where the two calculated curves cross they are smoothed into one.

In order to check the two approximate equa-

tions, one of the cases calculated by Baker was recalculated by means of the rigorous Eq. (27) developed in Section 4 and valid for the whole frequency range. This special case refers to a non-magnetic metal ($\mu=1$), an outer diameter of the cylinder of 10 cm ($a=5$), and a wall thickness of 1 millimeter ($b=4.9$). ρ is taken as 7×10^{-6} ohm cm ($=7 \times 10^4$ e.m.u.). Since the heat input is proportional to H_0^2 according to all the equations used, W/H_0^2 is calculated. A frequency range from 300 to 50,000 is covered.

$$\begin{aligned} A_r^2 + A_i^2 &= \kappa, \\ B_r^2 + B_i^2 &= \lambda, \\ A_r B_r + A_i B_i &= \mu, \\ A_r B_i - A_i B_r &= \nu. \end{aligned} \quad (26)$$

With these quantities, the heat input W becomes:

tions, one of the cases calculated by Baker was recalculated by means of the rigorous Eq. (27) developed in Section 4 and valid for the whole frequency range. This special case refers to a non-magnetic metal ($\mu=1$), an outer diameter of the cylinder of 10 cm ($a=5$), and a wall thickness of 1 millimeter ($b=4.9$). ρ is taken as 7×10^{-6} ohm cm ($=7 \times 10^4$ e.m.u.). Since the heat input is proportional to H_0^2 according to all the equations used, W/H_0^2 is calculated. A frequency range from 300 to 50,000 is covered.

Before giving the results of the complete calculation, the computation will be shown in detail for one particular frequency. A frequency very close to the crossing point of the two approximate curves is selected, i.e., 17,900. From Eq. (7) we then have $k=14.2$, which gives $ka=71.20$ and $kb=69.78$.

The first step is the evaluation of the sixteen Bessel functions involved for ka and kb as arguments. For arguments up to 10, the functions have been tabulated previously.⁵ They are sometimes found under the real and imaginary parts (Re and Im) of the respective J and $H^{(1)}$ functions. It is useful to remember the following identities:

$$\begin{aligned} ber\ x &= Re\ J_0(i^1 x), & ker\ x &= -\frac{\pi}{2} Im\ H_0^{(1)}(i^1 x); \\ bei\ x &= -Im\ J_0(i^1 x), & kei\ x &= -\frac{\pi}{2} Re\ H_0^{(1)}(i^1 x); \\ ber'\ x &= -Re\ i^1 J_1(i^1 x), & ker'\ x &= \frac{\pi}{2} Im\ i^1 H_1^{(1)}(i^1 x); \\ bei'\ x &= Im\ i^1 J_1(i^1 x), & kei'\ x &= \frac{\pi}{2} Re\ i^1 H_1^{(1)}(i^1 x). \end{aligned}$$

For arguments larger than 10, tables do not exist, but there are asymptotic formulas that permit the computation. The eight formulas are listed below. Let

$$F = e^{x/2}/(2\pi x)^{1/2} \quad F' = (\pi/(2x))^{1/2} e^{-x/2}$$

and

$$\varphi = x/(2)^{1/2} + \pi/8 \quad \psi = x/(2)^{1/2} - \pi/8.$$

Then:

$$\begin{aligned} \text{ber } x &= F[(1 + 0.088/x) \cos \psi \\ &\quad + (0.088/x + 0.070/x^2) \sin \psi], \\ \text{bei } x &= F[(1 + 0.088/x) \sin \psi \\ &\quad - (0.088/x + 0.070/x^2) \cos \psi], \\ \text{ber}' x &= F'[(1 - 0.265/x) \cos \varphi \\ &\quad - (0.265/x + 0.117/x^2) \sin \varphi], \\ \text{bei}' x &= F'[(1 - 0.265/x) \sin \varphi \\ &\quad - (0.265/x + 0.117/x^2) \cos \varphi], \\ \text{ker } x &= F'[(1 - 0.088/x) \cos \varphi \\ &\quad - (-0.088/x + 0.070/x^2) \sin \varphi], \\ \text{kei } x &= F'[-(1 - 0.088/x) \sin \varphi \\ &\quad - (-0.088/x + 0.070/x^2) \cos \varphi], \\ \text{ker}' x &= F'[-(1 + 0.265/x) \cos \psi \\ &\quad + (0.265/x - 0.117/x^2) \sin \psi], \\ \text{kei}' x &= F'[(1 + 0.265/x) \sin \psi \\ &\quad + (0.265/x - 0.117/x^2) \cos \psi]. \end{aligned}$$

Using these formulas, the following data are obtained for the particular frequency mentioned above:

$$ka = 71.20$$

$$\text{ber} = 0.3033 \times 10^{21} \quad \text{ker} = 0.01800 \times 10^{-21}$$

$$\text{bei} = -0.1070 \times 10^{21} \quad \text{kei} = -0.00920 \times 10^{-21}$$

$$\text{ber}' = 0.3068 \times 10^{21} \quad \text{ker}' = -0.01936 \times 10^{-21}$$

$$\text{bei}' = 0.1587 \times 10^{21} \quad \text{kei}' = -0.006165 \times 10^{-21}$$

$$kb = 69.78$$

$$\text{ber} = 0.03220 \times 10^{21} \quad \text{ker} = 0.04800 \times 10^{-21}$$

$$\text{bei} = -0.1246 \times 10^{21} \quad \text{kei} = 0.02828 \times 10^{-21}$$

$$\text{ber}' = 0.1106 \times 10^{21} \quad \text{ker}' = -0.01429 \times 10^{-21}$$

$$\text{bei}' = -0.06443 \times 10^{21} \quad \text{kei}' = -0.05414 \times 10^{-21}.$$

The next step is the evaluation of the eight quantities given by Eq. (15), by means of which A and B from Eqs. (16) and (17) are calculated. The four components shown in Eqs. (18) and (19)

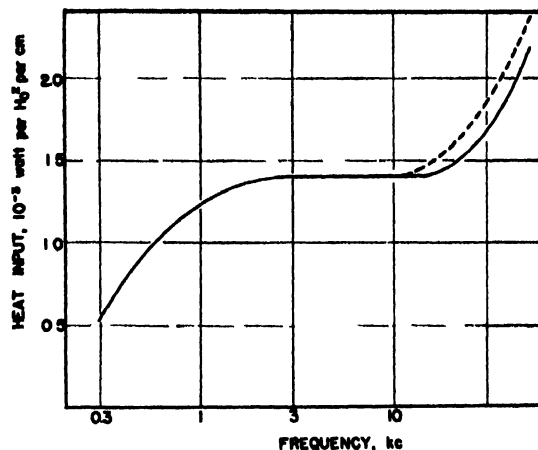


FIG. 1. Heat input as function of the frequency for a special case of a hollow cylinder. Solid line calculated from Eq. (27); dotted line calculated by R. M. Baker from approximate formulas.

are then in the present case:

$$A_r = 2.823 \times 10^{-21} H_0,$$

$$A_i = 0.8252 \times 10^{-21} H_0,$$

$$B_r = -1.609 \times 10^{21} H_0,$$

$$B_i = 0.8668 \times 10^{21} H_0.$$

These data permit evaluation of the four quantities shown in Eq. (26), and in this manner all magnitudes occurring in Eq. (27) are known. In the present case the result is

$$W = 1.43 \times 10^{-3} H_0^2$$

in watts/cm, if H_0 is measured in e.m.u.

The calculation was carried out for five selected frequencies and the result of the complete calculation, representing the heat input as a function of frequency, is shown in Fig. 1. The solid line is calculated from Eq. (27), the dotted line is Baker's curve. As can be seen, the two methods are strictly identical below the frequency for which the depth of penetration is equal to the wall thickness; for higher frequencies the approximate formula gives values slightly higher than those from the rigorous equation. For still higher frequencies the two curves merge again; it can be shown that for depths of penetration markedly less than the wall thickness Eq. (27) degenerates to the second of the approximate formulas given. The highest frequency for which computation was made is 4000 kc (not shown in Fig. 1); W , according to both equations, is 20.9 watts/cm.

The Tensile Properties of Several Special Steels and Certain Other Materials under Pressure

P. W. BRIDGMAN

The Physics Laboratories, Harvard University, Cambridge, Massachusetts

(Received November 16, 1945)

INTRODUCTION

IN a series of papers¹ the effect of hydrostatic pressure on the tensile properties of a large number of different steels has been investigated. The pressure effects on all these steels fell approximately into a one-dimensional pattern, the behavior being determined by a single parameter, the hardness or the absolute magnitude of the yield point at atmospheric pressure. In spite of the great ostensible variety of the steels, they were, nevertheless, restricted to a range of only a few percent in any of the conventional alloying elements, and to a hardness range below 50 Rockwell C. In the present investigation two different series of steels were used falling outside the previous limits. The first of these comprised four stainless steels running up to 18 percent Cr and 10 percent Ni; the second was a series of five different hardnesses of a carbon steel, running from 48 to 63 Rockwell C.

In anticipation of the results, it was found that two of the stainless steels fall markedly outside the previous pattern, but the others fell approximately within it.

In addition to the measurements on the steels, qualitative measurements were made on several other materials.

In the discussion, the new results of this paper and previous unclassified results are surveyed together.

APPARATUS AND METHODS

The apparatus and method for making the tensile tests have been fully described in the previous communications; only a brief characterization will be made here. The tensile specimens, of an over-all length of 1 inch and a

diameter varying about a mean of 0.1 inch, depending on the absolute strength, are completely surrounded by a fluid in which pressures up to 30,000 kg/cm² may be produced and maintained during stretching. The specimen is stretched by a forked yoke driven by the piston with which pressure is produced. The tension on the specimen is measured within the pressure chamber by the change of resistance of a "grid," in this way avoiding errors from the friction of the piston. Simultaneous measurements were made of hydrostatic pressure, tension on the specimen, and elongation of the specimen, for progressively increasing elongations. The elongation can be determined with adequate accuracy only in the range of plastic flow, after reaching the maximum load. The parameters singled out for determination from the complete data and presented in the following are: maximum load, load at fracture, reduction of area at fracture, and character of the fracture—all as functions of the pressure. In many experiments stretching was not pushed to fracture; for these the final loads were determined instead of the load at fracture.

In calculating the stresses from the data the change of geometrical configuration due to necking is so large that it was necessary to make allowance for the distribution of stress across the neck. This was done by methods which have been outlined in several papers.² The stress system at the neck of a tension specimen can be analyzed into the sum of three stress systems. The first of these is the hydrostatic pressure in the ambient liquid; the second is a tension along the axis of the specimen, *constant* across the section, called the "flow" stress in the following; the third is a variable hydrostatic tension (three equal components of principal stress) varying from zero at the outer surface of the neck to a maximum on the axis according to a formula given in detail in

¹ Eight Reports to the Watertown Arsenal by P. W. Bridgman, from March, 1943 to December 21, 1944, numbered No. WAL 111/7, No. WAL 111/7-1, etc., through No. WAL 111/7-7. See also P. W. Bridgman, *Rev. Mod. Phys.* 17, 3-14 (1945). There are also several reports by P. W. Bridgman to the N.D.R.C., on which the confidential classification had not yet been lifted at the time of writing.

² P. W. Bridgman, *Trans. A.S.M.E.* 32, 553-574 (1944); *Metals Technology*, Technical Publication No. 1792 of Am. Inst. Min. Met. Eng. (Dec. 1944) with supplement in April, 1945; *Wat. Ars. Report*, No. WAL 111/7-4.

the references. It follows that the "flow" stress is not equal to the average tension across the section (total tensile load divided by the actual area of the neck), but is less than it. The factor by which it is less will be called the "correction factor." In the following work the correction factor varied through an extreme range from 1.0 to 0.7.

THE MEASUREMENTS

1. Stainless Steels

Measurements were made on four different varieties of stainless steel. For these I am indebted to Dr. A. L. Feild of the Rustless Iron and Steel Corporation. The designations, compositions, and heat treatments are given in Table I.

The material was supplied in $\frac{1}{4}$ -inch rods. These had been drawn to approximately the final size, then annealed and pickled. After this they were given a final cold drawing pass with further reduction in cross-sectional area of 5 percent. The final hardnesses on the Rockwell *B* scale were 87, 93, 95, and 93, respectively, for H-26, H-50, H-29, and H-88.

Seven runs were made on each of these four steels at pressures ranging from atmospheric to 25,000 kg/cm². In four of these runs, including the run at atmospheric pressure, the test was carried to fracture, and in the three others it was stopped short of fracture. The purpose of the latter runs was to obtain points on the strain hardening curve comparatively free from the error introduced by the disturbing local flow which is very likely to occur in the moment of fracture.

These steels fall into two groups, of markedly different properties. At atmospheric pressure H-26 and H-50 break in the fashion normal for a steel of high ductility, with a reduction of area of about 78 percent. Necking starts, as is normal, at a strain of about 0.2. Under pressure these two steels show an increase of ductility so large that

TABLE I.

	C	Mn	P	S	Si	Cr	Ni	Mo	Cu
H-26	0.094	0.36	0.023	0.022	0.35	12.26	0.46	0.50	ND
H-50	0.067	0.50	0.020	0.030	0.31	17.49	0.41	ND	ND
H-29	0.058	0.70	0.030	0.013	0.85	18.51	8.95	ND	0.20
H-88	0.051	0.59	0.030	0.020	0.47	18.31	10.27	ND	0.20

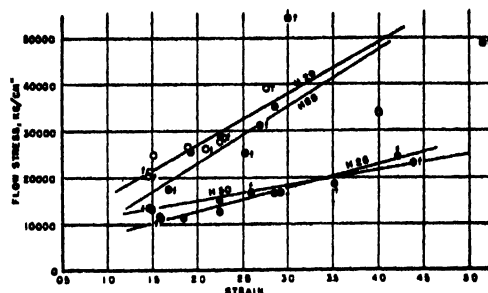


FIG. 1. The flow stress as a function of strain (natural) for four stainless steels. An *f* on a number of points indicates the occurrence of fracture.

the maximum pressure was not pushed beyond 19,000 kg/cm², the reduction of area at this pressure being so extreme that error is introduced by deviations from geometrical regularity because of finite grain size. The two other steels, H-29 and H-88, break with the same reduction of area at atmospheric pressure, but the character of the stress-strain curve is quite different, since the elongation is much greater and necking does not start until a strain of 0.65 is reached. Under pressure the increase of ductility is not nearly as large as for the two other steels, and pressure could be raised to 25,000 without enough reduction of area to introduce error from geometrical irregularity.

Because of the much greater reduction of area before necking starts, the contour of a tension specimen of H-29 or H-88 has a different shape from the other two or from the other steels investigated previously. Since the shape of the contour determines the correction for stress distribution across the neck, the shape of the contour, that is, the radius of curvature of the contour at the neck, was measured for all specimens of all four steels. The curve of a/R (radius of curvature of the contour at the neck divided by the radius of the cross section at the neck) against strain was normal for H-26 and H-50, but for H-29 and H-88 it lies below the normal curve in the initial stages, cutting the axis at a strain of 0.65 instead of at 0.2, and approaches coincidence with the normal curve at a strain of 3.5. The corrections for the stress distribution in H-29 and H-88 were determined by utilizing these special measurements of a/R .

The results are reproduced graphically. In Fig. 1 is shown the flow stress in terms of the

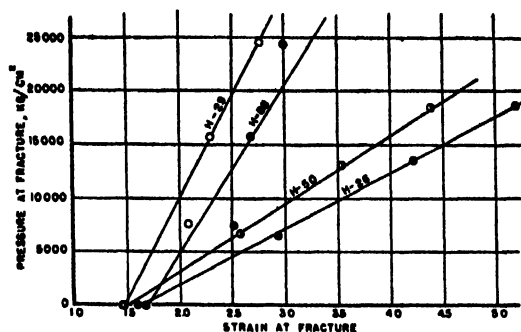


FIG. 2. The hydrostatic pressure at fracture as a function of the strain at fracture for four stainless steels.

strain; the concealed variable is the hydrostatic pressure, which in general is greater for the larger strains. An f written against some of the points indicates that the specimen was fractured during the test. In general the scattering of the points is somewhat greater for the fractures; this is due to distortion in the process of fracture, which may be considerable at the largest strains. Within experimental error the relation between flow stress and strain is linear in the present range. More accurate measurements under especially favorable conditions have shown that the curves are slightly concave toward the axis of strain.

In Fig. 2 is shown the pressure at fracture against the strain at fracture. In accordance with the great majority of previous results, the relation is assumed to be linear, and is so shown. The strain at fracture can be determined with some accuracy, however, and instances have already been found in which the relation is almost certainly not linear within experimental error. It is therefore not at all impossible that the relation for H-88 should be represented by a curve concave toward the pressure axis.

In Fig. 3 is shown the flow stress at the maximum load (that is, at the point where necking starts) as a function of the pressure of pulling. For all four steels this flow stress increases linearly with pressure, by some 3 percent per 10,000 kg/cm² for H-26 and H-50, and by some 13 percent for the same pressure increment for H-29 and H-88. The former rate, 3 percent, is somewhat less than for most steels previously investigated; the latter rate, 13 percent, is distinctly higher. Except for a factor constant with respect to pressure, the "flow

stress at maximum load" is the same as the technical "tensile strength." The factor to reduce from flow stress to tensile strength is 0.83 for the first two steels and 0.52 for the second two; this factor is determined by the strain at which necking starts, the "tensile strength" being calculated on the original cross section.

The fracture of all four stainless steels roughly follows the same pattern previously found. At atmospheric pressure the fracture is of the conventional "cup-cone" type. With increasing pressure the tensile part of the fracture (bottom of the cup) becomes progressively less in extent relative to the total cross section, and eventually disappears, the fracture becoming all shear above a certain pressure. In the lower part of the domain of completely shearing fracture the surface of shear is conical; at higher pressures this multiplies into a double cone, and at still higher pressures the shearing fracture shifts to a single distorted plane. These features are all present in the stainless steels. In Fig. 4 is shown the ratio of the area of the tensile part of the fracture to the total area of the fracture as a function of the hydrostatic pressure prevailing at the moment of fracture. The fracture becomes all shear at a lower pressure for the two steels H-26 and H-50 than for the two others, H-29 and H-88. The intermediate episode of shearing fracture on single or multiple conical surfaces is confined to a comparatively narrow pressure range, and was definitely observed for only one specimen, most of the shearing fractures being on a single distorted plane. The general character of the fracture was always more irregular and coarsely granular than for ordinary carbon steels.

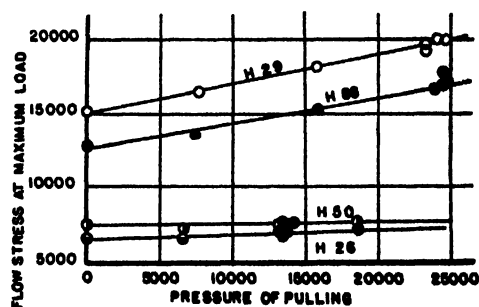


FIG. 3. The flow stress at the maximum load (i.e., at the beginning of necking) as a function of the prevailing hydrostatic pressure for four stainless steels.

2. Steels of Hardness Greater than Rockwell C 48

I am greatly indebted to Captain John H. Hollomon, who had the specimens prepared at the Watertown Arsenal under his direction. The specimens were all made from the same bar of high carbon steel. They were heat-treated by holding at 1525°F for 30 minutes and then quenching into oil at room temperature. They were then drawn in batches of ten to five different temperatures so chosen as to give median hardnesses of 48, 51, 56, 60, and 63.

These experiments followed a number of unsuccessful attempts to demonstrate an effect of hydrostatic pressure on the ductility of tool steels of hardnesses greater than 60. The preliminary attempts had encountered brittle failure

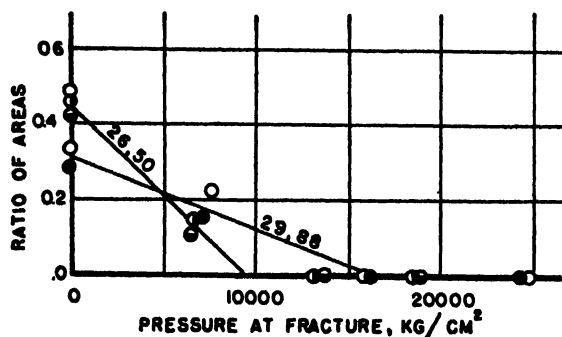


FIG. 4. The ratio of the area of the tensile part of the break (bottom of the cup) to the total area as a function of the pressure at which the break occurs for the four stainless steels. The two groups H-26, H-50 and H-29, H-88 are so nearly equal that a single line is drawn to represent the average behavior of the two.

at the shoulders of the test specimen because of stress concentration. After a number of trials the shape of the specimen shown in Fig. 5 was arrived at. The specimens were rough finished at the Arsenal to the shape shown and heat-treated. They were finished to final dimensions at Harvard by grinding; I am greatly indebted to the skill of Mr. Charles Chase for this operation. The minimum diameter of the finished specimen was varied somewhat with the absolute strength of the steel; it was so chosen as to demand approximately the same maximum load for all heat treatments. The hardness of each specimen was determined on the flats ground on the ends as indicated. From each batch of ten, the four or five were selected whose hardness was closest to the median. Since the primary object of these measurements was to determine how much ductility was imparted by pressure, these specimens were all pulled to fracture at pressures ranging from atmospheric to a maximum of 28,000 kg/cm². The strain at fracture at atmospheric pressure as a function of hardness is shown in Fig. 6.

The same sort of measurements were made on these as on the stainless steels.

In Fig. 7 is shown the flow stress at fracture as a function of the strain at fracture. The error from distortion in the act of fracture is much less for these hard steels than for the softer steels, and Fig. 7 therefore gives essentially the strain hardening curves. The points are somewhat more scattered than for the softer specimens. Again the relation is linear within the

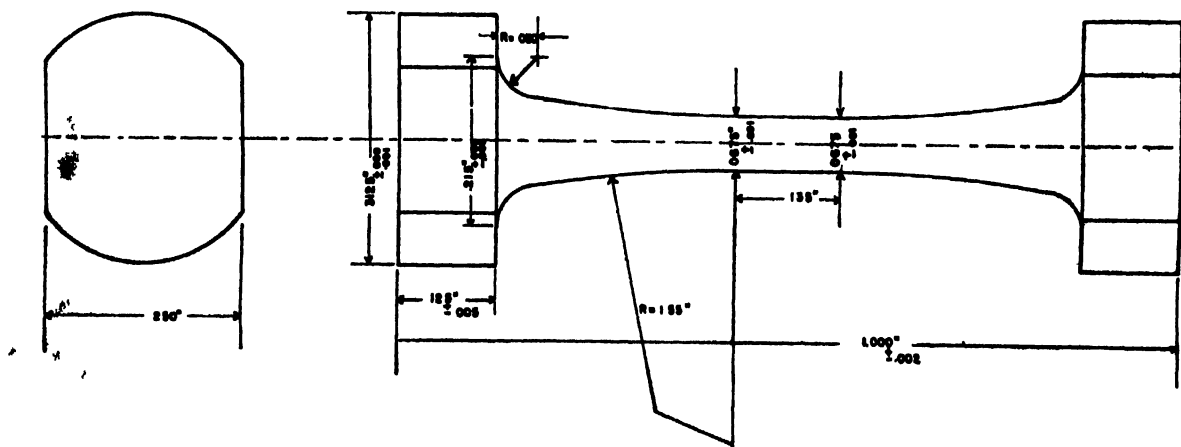


FIG. 5. The special shape of the tensile specimens of the hard steels. Diameters given provide for finish grinding.

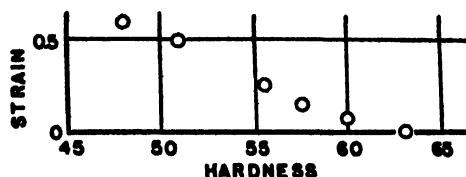


FIG. 6. The strain at fracture at atmospheric pressure of the hard steels as a function of the Rockwell C hardness.

limits of error. Two atmospheric points for hardness 63 lay distinctly out of line with the others and are not plotted; it is probable that there was some surviving stress concentration effect here. Only a single experiment under pressure was made on the 63 group, at 24,000 kg/cm². A slight but measurable ductility was found, the strain at fracture being 0.06. In view of the very small ductility at this pressure it did not seem worth while to try for measurements on this hardness at lower pressures.

In Fig. 8 is shown the hydrostatic pressure at fracture as a function of the strain at fracture. Again the relation seems to be linear within experimental error. The lines drawn in Fig. 8 are the best lines through the points of each group. Doubtless if a larger number of measurements had been made, lines would have been found more uniformly spaced in terms of the hardness. A vertical line in Fig. 8 means zero effect of pressure on ductility. The rapid decrease of pressure effect on ductility with increasing hardness in this hardness range is noteworthy.

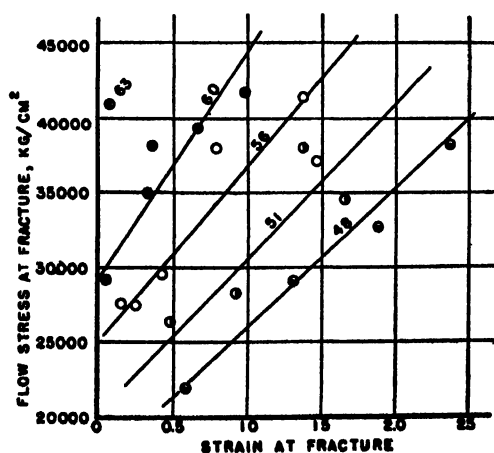


FIG. 7. The flow stress at fracture as a function of the strain at fracture for the hard steels. The figures on the lines indicate the average Rockwell C hardness of the group.

In Fig. 9 is shown the flow stress at maximum load as a function of the hydrostatic pressure at fracture. Since the hydrostatic pressure at fracture was only slightly higher than the hydrostatic pressure at the time of maximum load, this is practically a graph of flow stress at maximum load against the prevailing hydrostatic pressure. The factor for reducing from flow stress to tensile strength is 0.82 for all the specimens of Fig. 9. As in the case of Fig. 3, the

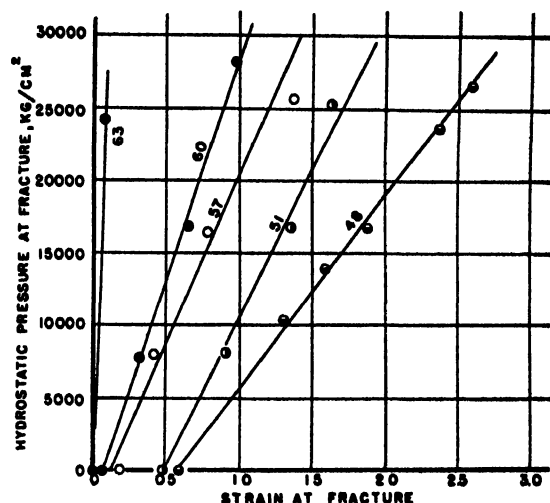


FIG. 8. The hydrostatic pressure at fracture as a function of the strain at fracture for the hard steels. The numbers on the lines indicate the average Rockwell C hardness of the group.

flow stress at maximum load increases somewhat with increasing pressure. The increase is greater on a percentage basis for the harder heat treatments.

In Fig. 10 is shown the ratio of the area of the tensile part of the fracture to the total fractured area as a function of the pressure. At atmospheric pressure the tensile part of the fracture occupies a progressively greater part of the total area as hardness increases, the fracture of the steel of hardness 63 being practically all on a single fine grained tensile plane except for a very narrow rim of shear around the edge. As usual, the tensile break disappears at high pressure. The pressure at which it disappears is practically the same for all the hardness groups except the hardest, which differs abruptly from the others. The multiple shearing fracture does not often appear in the harder steels, but at the lower end of the hardness

range is more common. The most common type of shearing break is on a single distorted plane; this is always highly burnished. In a couple of instances there were also fissures on longitudinal planes.

3. Tensile Properties under Pressure of Other Materials

In addition to the tests on the two series of steels just described, other sorts of material were examined, chiefly from a qualitative point of view and primarily to determine whether ductility is imparted to all sorts of material by pressure.

A number of experiments were made on gray cast iron. This was the ordinary gray cast iron of commerce, fine grained, and without obvious imperfections. Several experiments were first made in which the specimen was given the conventional shape of most of the previous specimens. These tests all gave negative results, the specimen fracturing under pressure with a clean tensile break and no evidence of plastic flow. Two specimens were then tried with the special shape shown in Fig. 5 to avoid any effects of stress concentration. These were both pulled at 29,000 kg/cm². One was exposed directly to the action of the pressure transmitting liquid, *isopentane*, whereas the other was copper plated in order to prevent penetration by the liquid into any pores that there might be. Both experiments gave the same completely negative result, clean fracture perpendicular to the axis of the speci-

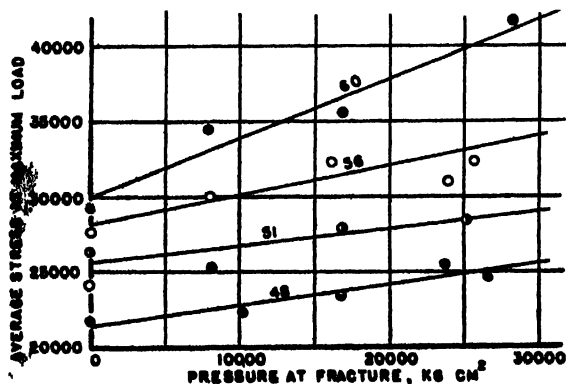


Fig. 9. The "average longitudinal stress," which in this case is the same as the flow stress, at the maximum load as a function of the hydrostatic pressure for the hard steels. The numbers on the lines indicate the average Rockwell C hardness of the group.

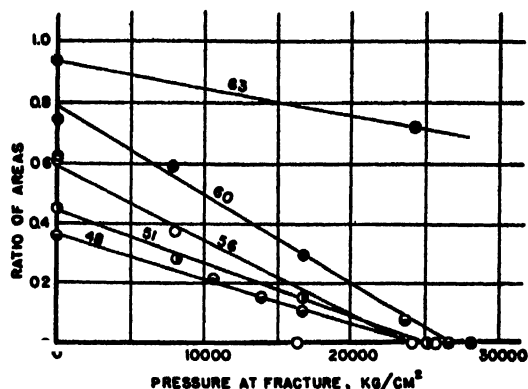


FIG. 10. The ratio of the area of the tensile part of the break (bottom of the cup) to the total area as a function of the pressure under which the break occurs for the hard steels. The numbers on the lines indicate the hardness group.

men, with no measurable reduction of diameter. It is perhaps natural to ascribe the negative results to the microscopic grains of pure graphite known to be present in this material.

For unity of presentation, certain results with other materials which have already been described elsewhere will be referred to again here.³ Tension tests under pressure on specimens of the conventional shape were made on aluminum, copper, bronze, and brass. The ductility of all these is markedly increased by pressure. In general it is not possible to push the strains to such extreme values without loss of geometrical regularity with these materials as with the steels. The cross section is likely to become dominated by a single crystal grain, so that the section becomes rectangular or square, with the result that the stress at the section becomes inhomogeneous and the determination of the stress-strain relation is thwarted. No serious attempt was made to find whether this effect was due to abnormally large grains in the virgin material or whether large grains were developed by the process of straining under pressure; my general impression in the case of aluminum and copper was that it was the latter.

Qualitative measurements were also made on a number of other materials that could not be readily fashioned into the conventional tensile specimens. These materials were made into slender rods, approximately 1 cm long and 0.8 mm in diameter, supported near the two ends

³ P. W. Bridgman, *Rev. Mod. Phys.* 17, 3-14 (1945).

and subjected to a push at the center by a blunt wedge. Stops so limited the travel of the wedge that ductile materials were bent to a maximum angle of 135°. The experiment consisted in exposing the rod to the bending action of the wedge under hydrostatic pressure and observing whether there was any measurable difference between the amount of bending in the neighborhood of the fracture (if there was a fracture) as compared with atmospheric pressure. The apparatus was first tried with rods of a 1.50 C tool steel, hardened glass hard and drawn to various temperatures. At 29,000 kg/cm² the undrawn specimen broke with no appreciable bending, but all the others received the whole 135° bend without fracture. Of these the specimen drawn at 200°C broke at atmospheric pressure with no appreciable bending and the one drawn at 275° broke after bending through a slight angle. Repetition of the same experiment at 25,000 kg/cm² showed a rather rapid falling off of the pressure effect, the specimen tempered at 200° now breaking after bending only 40°.

At 29,000 gray cast iron broke with no appreciable bending, thus checking the result by the other method.

Other materials tested by the same method were: antimony, tellurium, bismuth (all in the single crystal condition), borax glass, hard rubber, and limestone. A rod of bismuth was recovered after an exposure at 15,000 permanently bent without fracture to an angle of 80°; whether the difference between 80° and 135° was due to elastic recovery was not investigated. Antimony, arsenic, and tellurium might be expected to show some ductility in analogy with bismuth, but loading of these three materials at 29,000 resulted in clean transverse fractures with no trace of any bending before the occurrence of fracture. Hard rubber was recovered after exposure to similar conditions unfractured and with a set of 10°; this set entirely disappeared and the rod became straight again on standing 12 hours at room temperature at atmospheric pressure. The elastic limit of hard rubber is thus obviously markedly raised by pressure. The borax glass was from the same source as that which Birch and Dane⁴ had used

in measuring the effect of pressure on viscosity at elevated temperatures. This glass deforms at atmospheric pressure with measurable velocities at temperatures as low as 300°; it might therefore be expected to show some measurable deformability under pressure at room temperature. No evidence of any bending before fracture was found, however, at a pressure of 29,000. Doubtless time effects are of importance here; the load was applied at such a rate as to reach the maximum deformation in a time of the order of a minute.

Limestone was of special interest because it had already been extensively investigated in the experiments of Griggs and of Balsley.⁵ They found marked plasticity under compressive stresses in combination with hydrostatic pressure and also in *simple* tension, but much less ductility in bending. The pressures that could be reached in the present experiments were higher than those attained by Griggs and by Balsley, and the external conditions were somewhat more uniform, so that a positive result might perhaps be expected. The fracture was, however, completely brittle at pressures up to 29,000 kg/cm² when the test was performed in the regular way. It is known from the experiments of Birch and Bancroft⁶ and others that the behavior of rocks under pressure is in many cases qualitatively altered if the pressure transmitting fluid is prevented from having access to the interior of the specimen by jacketing it with a suitable covering. The experiments were therefore repeated, encasing the limestone in a heavy sheath of lead or copper or steel, and bending under 29,000. The lead and copper jackets were probably completely successful; the specimens were markedly bent, and there were no obvious flaws, but the tests were not entirely conclusive because the specimens were broken on attempting to remove the jackets at atmospheric pressure. In the case of the steel jacket, both the jacket and the specimen broke brittly with no obvious permanent deformation. The conclusion would therefore seem to be that the limestone is capable of receiving permanent deformation in tension

⁴ D. T. Griggs, *J. Geol.* 44, 544-571 (1936); *Am. Mineral.* 23, 28-33 (1938); *Bull. Geol. Soc. Am.* 51, 1002-1022 (1940). J. R. Balsley, *Trans. Am. Geophys. Union*, Part II, 519 (1941).

⁶ F. Birch and D. Bancroft, *J. Geol.* 46, 113-141 (1938).

without fracture at pressures as high as 29,000. The extension of this conclusion to ordinary limestone is complicated by the fact that there is a polymorphic transition in the neighborhood of 25,000 so that the plasticity demonstrated above is plasticity of another modification than that existing under normal conditions.

In the endeavor to acquire further understanding of the effect of pressure on the mechanical properties of glass hard steel, another type of purely qualitative experiment was performed which has been previously made for other materials. If a thin ring of hard rubber or glass is slipped over a tightly fitting solid steel core and then exposed to hydrostatic pressure in a medium in which ring and core are both totally immersed, the ring will be split, just as if it had been expanded by driving a tupp into it. The simplest explanation is that the steel core, because of its much lower cubic compressibility, forces the ring to be stretched beyond the dimensions that it would normally assume under the hydrostatic pressure. This suggests that, for brittle materials, the normal zero from which to figure the extension which will produce brittle tensile fracture is the configuration which the material would assume under the mean equivalent hydrostatic stress of the total stress system.

The corresponding experiment was made with a 1.50 C tool steel by fitting a thin ring of the steel in the glass hard condition over a Carboloy core. The two were ground together with a slight taper, and were pushed snugly together, so that the steel ring was under a slight initial tension. The ring with core was then exposed to a pressure of 29,000 kg/cm². On releasing pressure, not only was the steel ring unbroken, but it was still a tight fit on the core, showing that it had received no appreciable plastic set. If the elastic constants of Carboloy, as well as the cubic compressibility, are assumed to be about one-third the corresponding constants of steel, which has never been proved by direct experiment but which is not unplausible, it may be calculated that at the maximum pressure the steel ring was strained by about 2.5×10^{-3} beyond the dimensions which it would normally have had under a pure hydrostatic pressure of 29,000. This stretch is far beyond the elastic limit or fracture point at atmospheric pressure, and

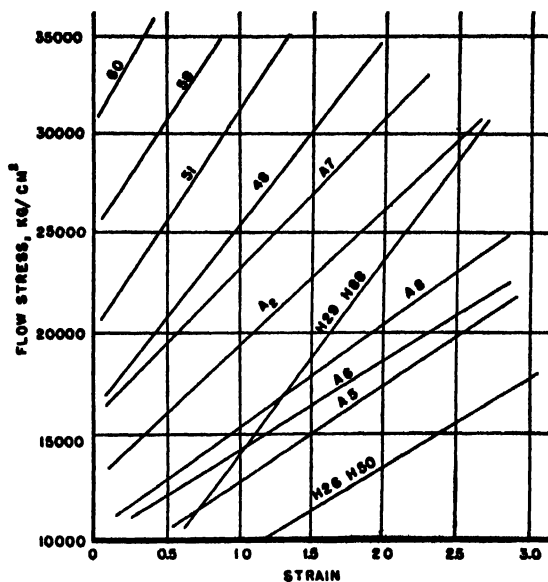


FIG. 11. The flow stress as a function of strain (strain-hardening curves) for all the steels of this paper and various 1045 steels, previously described in Watertown Arsenal reports.

would demand a tension of approximately 750,000 p.s.i. So large an effect of hydrostatic pressure on the elastic limit has not been found previously for the softer steels. It may be, therefore, although this conclusion should be confirmed by other methods, that the comparatively small effect of pressure on the ductility of very hard steels is compensated by a large effect on the elastic limit.

Special attention was given to an investigation of the effect of pressure on the tensile properties of Carboloy. The interest in this lies in the fact that Carboloy containers, immersed in a medium carrying a high pressure, successfully withstand an internal pressure of more than 100,000 kg/cm².⁷ The interiors of such containers are subject to large tensile strains and stresses, so that a large effect of pressure is indicated, Carboloy under normal conditions being brittle and with comparatively little strength in tension. The method used was the bending of slender rods, as before. No measurable plasticity in bending was imparted to a rod of grade 999 Carboloy by a pressure of 29,000 kg/cm². A special grid of greater sensitivity was then constructed and with this it was found that the bending force required to produce fracture at

⁷ P. W. Bridgman, J. App. Phys. 12, 461-469 (1941).

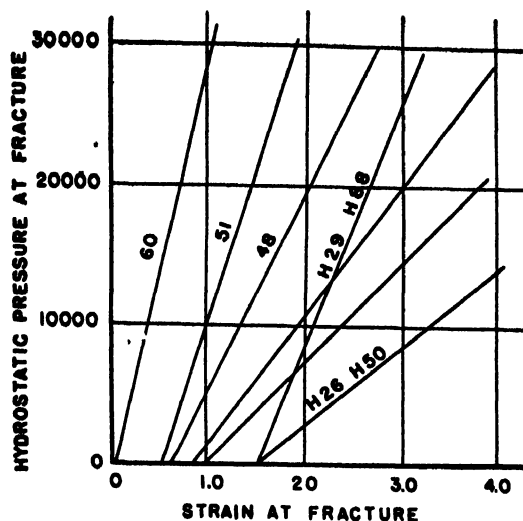


FIG. 12. Collected results for the steels of this paper and of a previous Arsenal report for the pressure at fracture as a function of the strain at fracture.

27,000 was at least threefold, and perhaps considerably more, greater than at atmospheric pressure. This very notable increase of strength must be an important factor in the functioning of the high pressure containers. The behavior is qualitatively the same as indicated in the last paragraph for hard steel.

DISCUSSION

In Fig. 11 are collected various results of the previous and the present investigations for flow stress *versus* strain for different grades of steel (that is, the strain hardening curves). The five lines designated by A2, A5, A6, A7, and A8 are

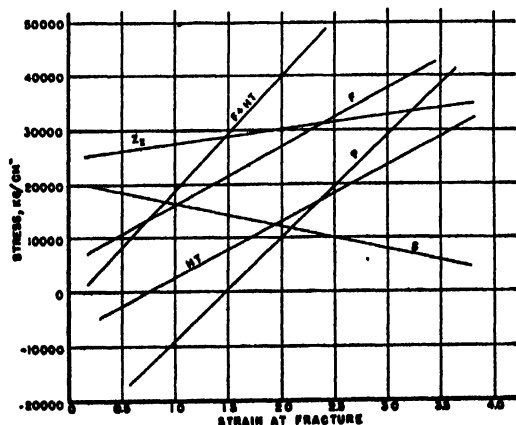


FIG. 13. Various significant stresses as a function of the strain at fracture for stainless steel H-29.

taken from the Watertown Arsenal reports, and refer to various heat treatments of a 1045 steel. The two groups of similar stainless steels are each averaged to give a single line. The lines of Fig. 11 form approximately a fan shaped family of non-intersecting lines, the slope of the line being higher the higher the absolute value of the stress, with the exception of the stainless steels H-29 and H-88. For these two steels the slope of the strain hardening line is higher than would be expected from the tensile strength at atmospheric pressure. That is, these two steels harden at an abnormally rapid rate.

Figure 12 shows a similar collection for different steels of the curves for pressure at fracture as a function of the strain at fracture, that is, the effect of pressure on ductility. The two undesignated lines are for the Watertown steels, which fall into two groups with respect to the property shown in Fig. 12. Again all the steels of this

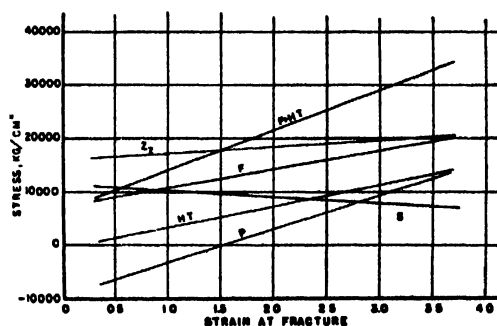


FIG. 14. Various significant stresses as a function of the strain at fracture for stainless steel H-50.

paper and two of the stainless steels fall into the same family of curves as the steels previously investigated, the increase of ductility under a given increment of hydrostatic pressure being greater the lower the absolute tensile strength. Again the two stainless steels H-29 and H-88 are exceptions. Their increase of ductility under hydrostatic pressure is abnormally low. Thus for these two steels a high rate of strain hardening with increase of strain goes with a low increase of ductility with increasing pressure.

The data of this paper bear on the question of the best criterion of rupture in the present range of conditions. In the fifth Arsenal report it was found that for the steels of those reports the mean hydrostatic tension (one-third the sum of

the three principal stresses, abbreviated in the following as S) was approximately constant at fracture. In Figs. 13-16 data are shown with respect to fracture for four representative steels of this paper, that is, stainless steels H-50 and H-29 and the groups of hardness 48 and 60. In these figures the lines of flow stress, F , and of hydrostatic pressure at the instant of fracture taken as positive, P , were first drawn as functions of strain. Next, from the line of flow stress was drawn the curve of "hydrostatic tension," HT , on the axis of the specimen. This hydrostatic tension is the second of the three components of the complete stress system described above. It was computed from the flow stress by multiplying by a somewhat complicated function of a/R which has been given in detail in the papers describing the corrections for stress distribution at the neck. This function of a/R may be expressed as one definite function of strain

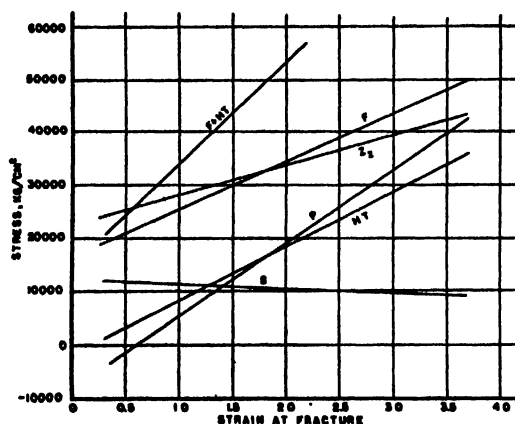


FIG. 15. Various significant stresses as a function of the strain at fracture for the specimens in the 48 hardness group.

for all normal steels, and as another function of strain for the two abnormal stainless steels. In both cases the conversion from a function of a/R to a function of strain was made with the use of empirical data. The function converting flow stress into "hydrostatic tension" is highly non-linear in the strain. It is a first empirical result that in the case of all steels, normal and abnormal alike, the curve for HT versus strain obtained with the use of this highly non-linear factor is linear within limits of error. It is a second empirical result that in all cases the line for HT

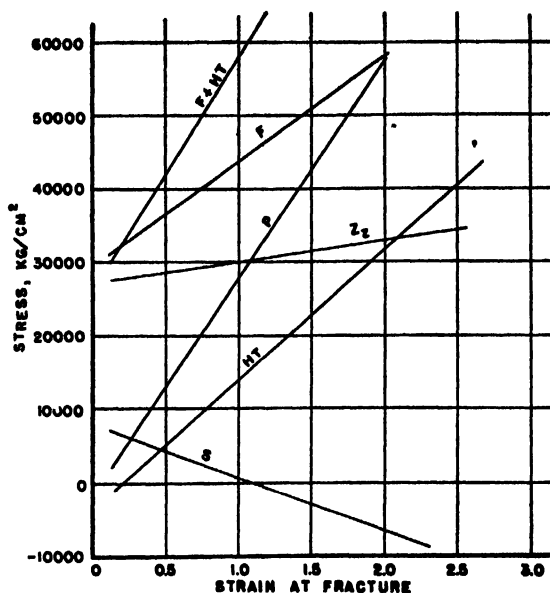


FIG. 16. Various significant stresses as a function of the strain at fracture for the specimens in the 60 hardness group.

rises approximately parallel to the line for F . Next, the two lines for F and HT were combined to give the line for $F+HT$. This is the axial stress at fracture that would be found on the axis of the specimen if there were no hydrostatic pressure. Inspection of the figures shows that in all cases it increases rapidly with strain. If the hydrostatic pressure is subtracted from $F+HT$ the actual stress Z , on the axis at the instant of fracture will be obtained. This line is also shown in the figures. Finally, the line for the mean hydrostatic tension ($S = \frac{1}{3}F+HT-P$) was constructed.

This method of representing the results is somewhat different from any that I have em-

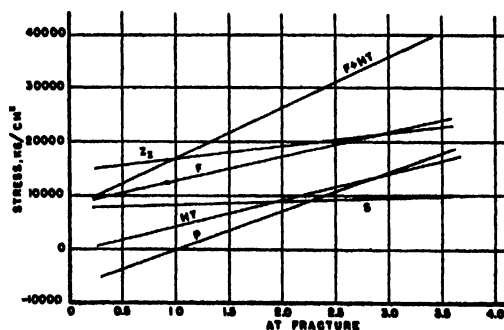


FIG. 17. Various significant stresses as a function of the strain at fracture for the steel A-5 of the fifth Arsenal report.

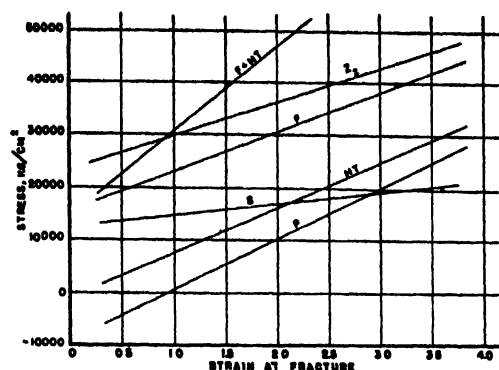


FIG. 18. Various significant stresses as a function of the strain at fracture for the steel A-7 of the fifth Arsenal report.

played before. Two of the steels previously examined in the Arsenal reports are therefore analyzed in the same way, and the results shown for A5 in Fig. 17 and for A7 in Fig. 18. These figures correspond exactly to the four figures just described for the steels of this paper.

In all six figures, 13 to 18, it is to be noted that the lines have physical significance for the conditions of the present experiments only for strains at which fracture occurs, that is, for strains to the right of the point where the line for P crosses the zero axis. To the left of this point fracture is formally indicated when the specimen is exposed to a hydrostatic tension instead of a pressure. What the actual physical behavior would be under such a tension is for the present conjectural.

A valid criterion for fracture will be represented by a horizontal line in the figures. Of the various lines in the figures it is obvious at a glance that the lines for Z , and S come nearest to meeting this requirement. It also appears that S does not stand up as well as a criterion for the steels of this paper as it did previously for the two Arsenal steels A5 and A7. In this connection too much weight should not be given to the apparently large failure for the steel of hardness 60, since the entire experimental range of strain was only 0.0 to 1.0 and the experimental error is therefore higher than for the others. The criterion that Z , is constant shows up as well on the average for the four steels of this paper as the criterion S . It is distinctly poorer for A5 and A7, however, so that on the whole it is not very

good. The physical significance of the criterion of constant Z , is that rupture occurs at a definite fiber stress, irrespective of the other stress components, and therefore irrespective of the component of elastic stretch in the same direction. This criterion has been used with some success by other investigators as a criterion of fracture in a limited range of conditions. However, in the wider range of conditions of this paper it would seem to have little physical significance, particularly when it is considered that under pressure the fracture loses its tensile character and becomes entirely shearing. The criterion definitely fails for such phenomena as the "pinching-off" effect.⁸ The physical significance of the criterion that S be constant is that fracture occurs when the average volume distension reaches a critical value. There would seem to be some physical plausibility in such a criterion, but the data of this paper, as well as considerations which I have already advanced in other papers, show that it cannot be the sole determining factor in fracture. However, it seems to me that this must contain an important aspect of the mechanism responsible for fracture.

Of the various criteria which have been suggested for fracture, it is obvious that no criterion in terms of strain only can possibly be valid under present conditions in view of the enormous strains that can be reached under high pressures without fracture. Of the simpler criteria in terms of stress, it is obvious that the maximum shearing stress criterion cannot be valid. The maximum shearing stress is simply $\frac{1}{2}F$, and inspection of Figs. 13-18 shows that F and therefore $\frac{1}{2}F$ in all cases rises rapidly with strain, so that this criterion cannot be general. A physical argument could be made for the condition that $F + HT$ should be constant at fracture, since this has the physical significance of maximum tensile fiber stress calculated from a state of uniform hydrostatic pressure as the zero. This criterion however, evidently fails by a wider margin than any other indicated by the figures. Or, since the total component of hydrostatic stress on the axis, where fracture begins, is $HT - P$, a similar argument could be made for the reasonableness of the criterion that F should be constant at

⁸ P. W. Bridgman, Phil. Mag. 63-80 (July, 1912).

fracture. This criterion is also evidently wide of the mark.

It is probably worth while calling attention to the fact that for all the steels the value of S at the lowest strain at which fracture occurs is approximately the same, namely $10,000 \text{ kg/cm}^2$.

SUMMARY

The tensile properties of the two stainless steels with low nickel content fall into the same pattern as the carbon or low alloy steels previously investigated. Two stainless steels with 9 and 10 percent nickel content fall outside the previous pattern in that necking starts at an abnormally high elongation, they strain harden at an abnormally rapid rate, and the increase of ductility under pressure is abnormally low. The carbon steels with Rockwell C hardness in the range 48 to 63 all fall in the previous pattern, the rate of strain hardening being greater and the effect of pressure on ductility being less the greater the hardness. Even the steel of hardness 63 shows a measurable ductility under a hydrostatic pressure of $24,000 \text{ kg/cm}^2$.

A search is made for a possible criterion of rupture by plotting various relevant stresses against the strain at fracture. No criterion emerges good under all conditions. The constancy of the mean hydrostatic tension (one-third the sum of the three principal stress components) remains the best criterion over the whole range of conditions, but in certain circumstances it may show considerable variation, and its margin of superiority over the criterion that the total fiber stress in the direction of fracture, Z , be constant at fracture, is not impressive. The maximum shearing stress criterion for fracture fails by a wide margin. No criterion of fracture in terms of strain components only can possibly be valid. Certain other plausible criteria on the stress components are also found not to hold.

In addition to the generalizations obtained from previous work, namely that the strain hardening curves are linear in the strain and that the ductility is a linear function of the hydrostatic pressure, two new empirical results emerge from the present method of plotting. These are:

first, that the component of hydrostatic tension on the axis of the specimen arising from necking is a linear function of the strain at fracture and, second, that the line of hydrostatic tension *versus* strain is approximately parallel to the line of flow stress *versus* strain.

With regard to other metals than steel, aluminum, copper, bronze, and brass show an even greater effect of pressure on ductility than do the mild steels. These four metals were tested in the form of conventional tension specimens. The tensile properties of a number of other materials were tested by bending slender rods under pressure. Results were obtained in this way on hardened tool steels in various tempers in line with those obtained with the conventional tension specimens. Antimony, arsenic, tellurium, borax glass, cast iron, limestone (uncovered), and Carboloy showed no measurable plastic deformation before breaking by bending under hydrostatic pressure. Bismuth shows large plastic effects. Hard rubber acquires great plasticity and also shows prominent time effects. Limestone, when covered, will take permanent plastic deformation in bending under hydrostatic pressure, but the interpretation of the results is obscured by a polymorphic transition which takes place under pressure. Although Carboloy showed no plasticity in bending, it showed a bending strength under $29,000 \text{ kg/cm}^2$ more than three times greater than that at atmospheric. Another type of experiment pointed to a similar very large increase of the elastic distortion which glass hard steel will support without fracture in a medium under hydrostatic pressure. In general comment on the bending experiments, it would seem that because of the complicated nature of the conditions the results are not susceptible of as straightforward an interpretation as are the conventional tension tests. This is pointed by the fact that carboloy will show plasticity in simple compression under pressure although it shows no plasticity in bending. The same thing is indicated by the marked difference of plasticity of limestone in simple tension and in bending.

I am much indebted to Mr. L. H. Abbot for making the readings and manipulating the pressure apparatus.

Letter to the Editor

Letter to the Editor

GEORGE G. BRUCK
157 South Harrison Street, East Orange, New Jersey
December 22, 1945

IN 1937 Woodyard published in the *Proceedings of the I.R.E.* a paper in which he described the use of an oscillator as frequency discriminator. Recently the growing interest in circuit arrangements permitting frequency discrimination without amplitude sensitivity has brought about such developments as the "Ratio Detector."

The circuit described here is essentially an oscillator which falls in step with an incoming signal. The differential plate-current of the two halves of the tube is the discriminator output (Fig. 1).

This circuit lends itself to many other applications like f.i. frequency modulation, frequency stabilization, frequency modulation phonograph pick-up, etc.

A more complete paper on the theory and experimental results with this circuit will be published shortly.

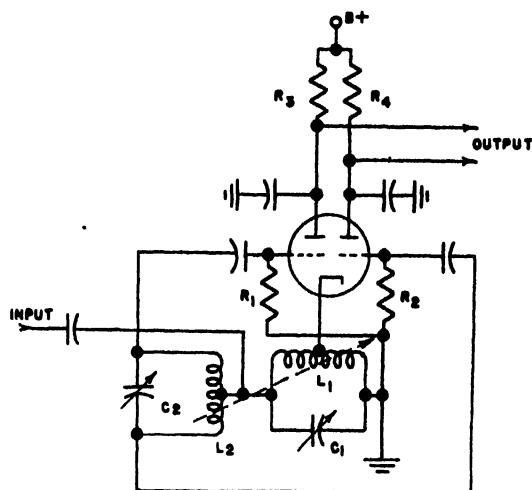


FIG. 1. F-m discriminator.

Here and There

New Appointments

Philip Nolan, who has spent the past three and a half years in fire control optics and infra-red research and development in the Navy's Bureau of Ordnance, recently joined the staff of the Farrand Optical Company, Inc., of New York as its chief physicist.

School and Society reports that G. W. Fox, professor of physics, Iowa State College of Agriculture and Mechanic Arts (Ames), has been appointed technical adviser to General Douglas MacArthur. Dr. Fox will spend about three months in Japan making recommendations as "to which scientific activities should be encouraged . . . and which, if any, discouraged."

After more than three years of war research with the Office of Scientific Research and Development and the U. S. Navy Bureau of Ordnance, George M. L. Sommerman has become associate professor of electrical engineering at The Technological Institute of Northwestern University.

Three new appointments were recently announced by the National Bureau of Standards. Hugh L. Dryden is now Assistant Director of the Bureau, Ferdinand G. Brickwedde is chief of the Heat and Power Division, and Robert Simha has become a member of the Division of Organic and Fibrous Materials with special responsibility for the coordination of fundamental research on the molecular properties of high polymers.

Necrology

Dr. Parke Benjamin Fraim, associate professor of physics at the Polytechnic Institute of Brooklyn, died January 13 at the age of 59.

Professor Harry Bateman, mathematician at California Institute of Technology, died in Utah on January 21 while on the way to New York, where he was to have been honored for outstanding work with Albert Einstein and other leading scientists. Professor Bateman was to be made a Fellow of the Institute of Aeronautical Sciences and receive a medal for his contributions to aeronautics. He had held full professorships in three fields—mathematics, physics, and aeronautics—and was the author of many standard higher mathematics texts. At his death he was 63 years of age.

Physicist Honored in Chicago

Dr. Glenn T. Seaborg, co-discoverer of plutonium and elements 95 and 96, who is "on loan" to the Metallurgical Laboratory of the University of Chicago, has been named the outstanding young man of Chicago for 1945 by the Chicago Junior Association of Commerce. Dr. Seaborg has been on leave from the University of California for the past four years. He discovered plutonium in 1940. The announcement of his discovery, with three co-workers, of elements 95 and 96 was made last November. Dr. Seaborg, who is 33, has been in charge of the research at the University of Chicago for separating plutonium in amounts and purity required for war purposes. He has also been primarily responsible for the chemical separation procedures which were used in connection with the manufacture of plutonium at Clinton, Tennessee, and Hanford, Washington.

Technical Translations

The R. T. P.'s listed below are available from the Durand Reprinting Committee in care of California Institute of Technology, Pasadena 4, California.

These were selected for reprinting from the British Ministry of Aircraft Production R. T. P. List No. 122.

- 2478 Hohenemker *Flow Tests on Steel Tubes under Combined Tension and Torsion.* (Z.A.M.M., Vol. II, No. 1, February, 1931, pp. 15-19.) (5 pages)
- 2480 Stablein, W. *Measurement of the Characteristics of Frequency-Modulated Oscillations.* (Elektrische Nachrichten-Technik., Vol. 4, April, 1943, No. 20, pp. 102-111.) (12 pages)
- 2481 Heuvelink, H. *The Testing of Graduations of Theodolites and Universal Instruments.* (Ztschr. f. Instrumentenkunde, Vol. 45, February, 1925, pp. 70-84.) (15 pages)
- 2492 Deuker, E. A. *On the Stability of Elastic Shells Pt. I.* (Z.A.M.M., Vol. 23, April, 1943, No. 2, pp. 81-100.) (26 pages)
- 2493 Deuker, E. A. *The Stability of Elastic Shells Pt. II.* (Z.A.M.M., Vol. 23, June, 1943, No. 3, pp. 169-179.) (13 pages)
- 2494 Pischinger, A. *Oscillations in Gas Columns.* (Forschung, Vol. 6, No. 6, November-December, 1935, pp. 273-280.) (12 pages)
- 2495 Schmidt, F. R. *The Induction and Discharge Process and Limitations of Poppet Valve Operation in a Four-Stroke Aero Engine.* (Luftfahrtforschung, Vol. 16, No. 5, 20.5.1939, pp. 251-275.) (45 pages)
- 2497 Volkersen, O. *Rivet Load Distribution in Lap Joints of Constant Cross-Section under Tension.* (Heinkel Werke, Rostock L. F. F., Vol. 15, No. 3, 20.1.38, pp. 41-47.) (14 pages)
- 2498 Jegorow, G. *Measurement of Direction Velocity and Pressure in a Three Dimensional Current.* (Luftfahrtforschung, Vol. 19, No. 9, 20.10.42, pp. 326-330.) (10 pages)
- 2499 Hiedemann, E. *Physical-Chemical Effects of Supersonic Waves.* (Archiv f. d. Eisenhüttenwesen, Vol. 12, No. 4, October, 1938, pp. 185-193.) (17 pages)
- 2500 Lutz, O. *Graphical Determination of Wall Temperature during Heat Transfer.* (Forschung, Vol. 6, No. 5, 1935, pp. 240-244.) (7 pages)
- 2501 Chwalla, E. *The Theory of One-Sided Web Stiffeners.* (Der Stahlbau, Suppl. to Bautechnik, Vol. 10, No. 10, May 7, 1937, pp. 73-76.) (10 pages)

Bulletin of the Atomic Scientists of Chicago

The Atomic Scientists of Chicago have published several issues of the first volume of their Bulletin, an eight-page

printed leaflet. Thus far articles have appeared under the following headings:

Technical Feasibility of Atomic Energy Controls
General Groves on the Future of the Atomic Bomb Projects
The Federation of Atomic Scientists
Denver Conference on Atomic Energy
UNO Faces the Problem of Atomic Energy Control
Structure of the UNO
British Public Opinion and the Atomic Bomb Menace
The Atomic Scientists Back the McMahon Bill
The Atomic Scientists and the Federation of American Scientists
The Distribution of Uranium in Nature
UNO Creates Atomic Energy Commission

In addition, the Bulletin always carries several columns under the heading "Congressional News."

The Bulletin has been featuring a series of articles on "Control Problems." Thus far it has included the following: general survey of the control problem, control of uranium mining, and control of a gaseous diffusion plant. The Bulletin is published at 1126 East 59th Street, Chicago 37, Illinois.

New Firm of Consulting Physicists

Paul Rosenberg Associates, consulting physicists, have announced the opening of offices in New York City. Work will be done in radar, ultrasonics, television, optics, nuclear physics, and general physics. The head of the firm, Paul Rosenberg, has recently completed four years' work at the Radiation Laboratory of the National Defense Research Committee at the Massachusetts Institute of Technology.

Dr. Meitner Arrives

Dr. Lisa Meitner, who is credited with establishing the mathematical data and equations in the breakdown of the uranium atom, arrived in New York from Stockholm on January 25. She came to the United States to join the faculty of the Catholic University of America in Washington.

New Booklets

Rheology Bulletin for December 1945, publication of The Society of Rheology, has the following table of contents:

Bibliography of Books on Viscosity, Viscometry and Rheology
Society of Rheology Annual Meeting Abstracts
Rheology Progress Abstracts (covering apparatus and methods, cellulose, clays, fibers, foams, general, ink, paints, petroleum products, plastics, polymers, proteins, theory)

General Radio Company, 275 Massachusetts Avenue, Boston 39, Massachusetts, has issued a 68-page *Postwar Supplement* to its *Catalog K*, now in the fourth edition. The Supplement brings all editions completely up to date.

Westinghouse Newsfront for January 1946, published by Westinghouse Electric Corporation, Pittsburgh 30, Pennsylvania, features articles on devices used in radar, research into the causes of rust and corrosion, industrial use of high frequency "cooking," and the value of electricity in increasing production, efficiency, and comfort on American farms.

Jarrell-Ash Company of 165 Newbury Street, Boston 16, Massachusetts, has issued a 16-page catalog of spectrographic equipment for research in spectroscopy and for spectrochemical analysis in metallurgy, mineralogy, agronomy, chemistry, medicine, and other fields.

Journal of Applied Physics

Volume 17, Number 4

April, 1946

A New Electrolytic Selenium Photo-Cell¹

A. VON HIPPEL, J. H. SCHULMAN, AND E. S. RITTNER

Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received December 20, 1945)

A new electrolytic selenium photo-cell is described and its characteristics given as a function of operating voltage, temperature, external resistance, and time. Data are included indicative of its response to unmodulated, modulated, and monochromatic light; comparison is made with cells of similar type or possessing similar characteristics; and a method of compensating for the dark current of photo-cells is shown.

An electronic circuit is also described which can greatly improve the voltage output of the photo-element. It is generally applicable to photo-conductive and photo-voltaic cells.

INTRODUCTION

A PROCESS of electroplating metallic selenium has recently been developed in this laboratory.² In the course of that research it was found that electrolytic selenium photo-cells of excellent sensitivity and relatively high impedance could be produced. These characteristics made it desirable to investigate in detail the performance of the new photo-elements in order to determine their usefulness in war equipment. The results of this study are presented below.

DESCRIPTION OF PHOTO-CELL

The electrolytic photo-cell consists of a metal electrode, completely coated with metallic selenium, immersed in an aqueous solution of an electrolyte, preferably selenium dioxide, together with an auxiliary electrode of a noble metal-like platinum, or a second selenium-coated electrode. The particular model of the glass-enclosed cell

employed in the major part of our experimental work is illustrated in Fig. 1. On the application of a small external d.c. voltage making the auxiliary electrode the anode, the cell becomes immediately sensitive to light incident upon the selenium cathode. The cell so employed as a photo-conductor exhibits essentially instantaneous and linear response to light. With a selenium dioxide electrolyte and an applied e.m.f. of 2 v the cell possesses a short circuit sensitivity of the order of 1000 μ amp./lumen and a dark current of 0.2–2.0 μ amp. (total cathode area = 2 cm²).

The most important factors controlling the quality of the cell are (1) the composition of the electrolytic solution and (2) the nature of the selenium coating on the cathode. The first may impose a limitation upon the useful life of the photo-cell; the second affects principally the short-circuit current and the voltage sensitivity.

THE ELECTROLYTE

Photo-sensitivity has been observed in cells containing a large number of different aqueous electrolytic solutions (Table I). In all of these

¹ This research was carried out under contract OEMsr 561 with Division 16, Section 16.4, of the National Defense Research Committee, and finished by June 10, 1943. Publication has been delayed because of the war.

² To be published elsewhere.

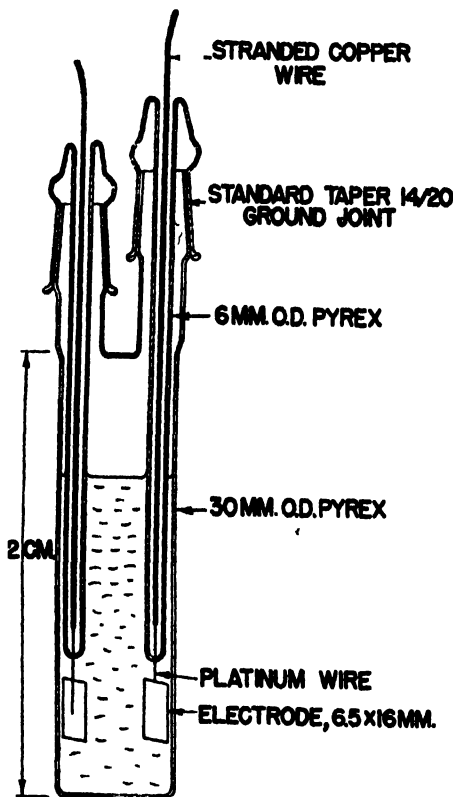
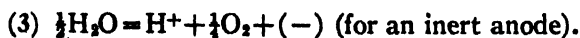
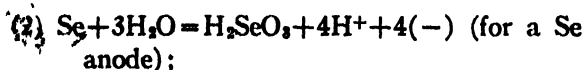
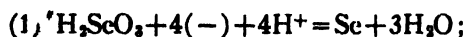


FIG. 1. The electrolytic selenium photo-cell.

cases there appears to be an optimum voltage, below or above which either the sensitivity decreases or the short circuit current output *vs.* illumination deviates from linearity. Employment of these solutions, however, results in the disadvantage that the light sensitive selenium surface is gradually destroyed on illuminating the cell either because of the evolution of H_2Se or of the dissolving of selenium by the locally high concentration of OH^- produced at the cathode in the case of salt electrolytes.

On the other hand, if selenious acid is used as electrolyte, the sensitive cathode is not destroyed, the electrode reactions which occur upon illumination being (1) at the cathode and (2) or (3) at the anode:



For this reason our detailed investigation of the

properties of the photo-cell was confined to cells containing selenious acid solutions.

The selenium dioxide employed was of the best grade commercially available, and its state of high purity was verified by spectroscopic analysis. The major impurity was Te, present in a concentration of 14 parts per million. All other metals found in the sample were present in concentration less than 2 parts per million. Despite this fact samples of selenium dioxide obtained from different companies and even different batches from the same manufacturer gave rise to cells of widely varying quality. Cells made with unsatisfactory samples of SeO_2 , for example, displayed abnormally high dark currents, or low sensitivities, or non-linearity and sluggishness of response. These differences in behavior would seem to be attributable to traces of impurities, undetectable spectroscopically, in the various lots of SeO_2 employed, as no correlation could be drawn between the spectroscopic analyses and the results obtained with varying SeO_2 lots. Moreover, attempts to purify unsuitable samples have failed to improve the characteristics of cells made from them, yielding no clue as to the nature of these presumed impurities. It was therefore necessary to select for further studies a single lot of selenium dioxide which gave rise to cells of desirable characteristics.³ Even with cells made from the same satisfactory solution and with electrodes plated simultaneously in the same plating bath, statistical fluctuations in cell characteristics were observed. Hence all data given below, except where noted, are representative of the average behavior of a minimum of five cells.

TABLE I. Photo-sensitivity as function of electrolyte.

Composition of solution	Short-circuit sensitivity, $\mu\text{amp./lumen}$	Dark current, $\mu\text{amp.}$	Optimum operating voltage (for Se anode)
0.1M HCl	510	0.1	1.0
0.1M H_2PO_4	610	0.2	1.0
0.1M H_2SO_4	620	0.6	1.0
9.0M H_2SO_4	680	2.	1.0
0.3M H_2SO_4	540	4.	0.55
0.1M HAc	500	2.	0.9
8.7M HAc	500	2.5	1.0
0.1M Na_2SO_4	630	0.5	1.0
0.1M $MgSO_4$	800	0.3	1.0

³ This SeO_2 lot was obtained from the American Smelting & Refining Company.

The concentration of the selenious acid may be varied within wide limits without materially affecting the cell's characteristics. Survey experiments showed that these limits extend from at least 4 to 60 weight percent SeO_2 . Since the eutectic of the system $\text{SeO}_2\text{--H}_2\text{O}$ lies at 57 weight percent SeO_2 and -23° , this composition seems preferable if the cell is ever to be subjected to low temperatures. Most of the data were obtained on cells whose electrolyte concentrations were close to the eutectic composition.

THE CATHODE

Although there are several techniques available for coating a metal with metallic selenium, we have confined our attention to the electroplating method as being the simplest and the most satisfactory for our purposes.

Metal electrodes can be electroplated with metallic selenium from a number of plating baths of which two were used in the present work: (1) the "standard" plating bath developed in this laboratory and (2) selenious acid solution (i.e., the cell solution itself) at elevated temperatures, 100° . With the first of these baths, metallic selenium may be plated to large thicknesses with high current efficiency, remarkable speed, and at low voltage. The second operates with low current efficiency owing to hydrogen evolution at the cathode. Strong polarization occurs, voltages of the order of 30 volts being required to pass currents of the order of 30 $\mu\text{amp.}$ after plating has proceeded for only a few minutes. This latter bath is thus unsuited for the production of thick plates.

The plates obtained from the two baths differ in appearance to the unaided eye and under the microscope. Bath (1) gives plates of about 4 percent average reflectivity, the Se crystals presenting the aspect of a grass-like growth with the C-axis oriented preferentially normal to the surface of the electrode. Bath (2) gives a much finer grained plate of about 25 percent average reflectivity with no sign of crystal orientation. Plates from bath (1) are easily wetted by water or SeO_2 solution, those from bath (2) shed liquid. Corresponding to the above-mentioned reflectivities, the former plates are very black in appearance, the latter a light gray.

When Pt electrodes plated in bath (1) are formed into cells (2 v applied), the initially high values of photo-sensitivity and dark current decrease rapidly with time.⁴ An equilibrium dark current of 0.2–2.0 $\mu\text{amp.}$ is attained within ten minutes to two days while a steady sensitivity of about 1000 $\mu\text{amp./lumen}$ is reached within a few days. This sensitivity is maintained over long periods of time, whether the cell is kept on voltage in the dark or stored on voltage under intense light.

With Pt electrodes plated in bath (2) for 5–10 minutes the situation is similar except for the fact that the equilibrium sensitivities are of the order of 500 $\mu\text{amp./lumen}$ and are attained only after one to two weeks. By exposing such cells to strong illumination for several hours the original grey color of the cathode is altered to a deep black, of about 4 percent reflectivity. Microscopic examination of the cathode reveals that an orientated structure similar to that produced in bath (1) has been formed. The change in character of the cathode is accompanied by an increase in sensitivity to 1000 $\mu\text{amp./lumen}$, and this sensitivity is thereafter maintained.

Other metals, besides Pt, which have been found suitable for use as bases for the Se cathode include Fe, Ni, Mo, and stainless steel. Cathodes plated in bath (1) on these metals gave cells with initially low sensitivities which rose rapidly to their equilibrium values within 1–2 days. With Fe and Ni the final sensitivities were 800–900 $\mu\text{amp./lumen}$; with Mo and stainless steel the final sensitivities were 1000 $\mu\text{amp./lumen}$.

All data given below refer to a total cathode area of 2 cm^2 whenever the dark current is concerned but to a light sensitive area of 1 cm^2 . The back side of the cathode, which is not illuminated, can be coated, if desired, with an inert insulating layer such as polystyrene to reduce the dark current.

THE ANODE

The anode of the photo-cell may consist of any inert metal or it may be a duplicate of the

⁴ The change in characteristics with time is so rapid initially that no attempt was made to measure photo-sensitivity until ten minutes after the application of voltage. The sensitivity of different cells at this stage varied between 1200 and 2400 $\mu\text{amp./lumen}$.

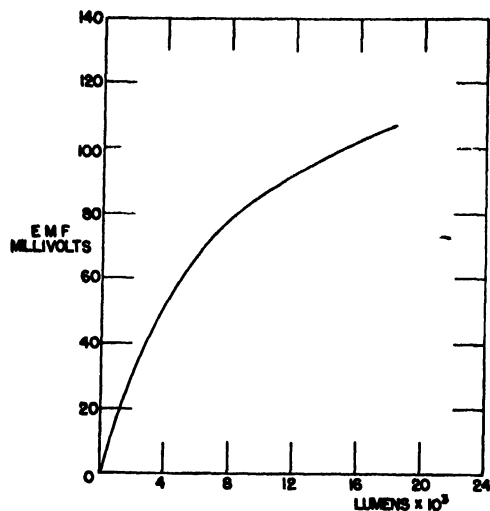


FIG. 2. Open circuit e.m.f. vs. illumination. (Illuminated electrode more positive)

cathode. With an inert metal anode, oxygen is evolved; with a Se-plated anode, selenium dissolves forming selenious acid—the reverse of the cathode reaction. If the cell is to be hermetically sealed, the evolution of gas is undesirable; hence, Se-plated anodes are to be preferred. In this case, however, the selenium on the anode must be of sufficient thickness to last for the lifetime of the cell. For this reason plating of the anode is necessarily done in bath (1) from which thicknesses sufficient for several years of operation may easily be plated. The further possibility exists in cells supplied with Se-plated anodes and cathodes of reversing the polarity at intervals, in which case the cell might be expected to have an indefinitely long lifetime.

Cells with selenium anodes possess slightly higher sensitivities and dark currents at a given applied voltage than cells with platinum anodes, probably owing to the higher back e.m.f. of the cell in the latter instance.

Se electrodes made on a Pt, Mo, or stainless steel base are suitable for anodes; those made on a Ni or Fe base are apparently too porous for such use, the solution becoming contaminated with oxidation products of Fe or Ni which destroy the photo-cell.

OPERATING VOLTAGE

The cell can be stored on open circuit without deterioration. The magnitude of the open circuit

potential generated, when one of two selenium electrodes in selenious acid solution is illuminated, varies greatly with individual electrodes, but the general shape of the cell e.m.f. vs. illumination curve is indicated in Fig. 2. Without applied potential the photo-current observed on illuminating a short-circuited cell quickly decays. This is likewise true of cells operated with applied potentials of less than about 0.5 volt, although the magnitude of the photo-current first observed increases with applied voltage. With further progressive increase of operating voltage the short-circuit sensitivity can be measured and typical data obtained on one cell with a Se-anode are shown in Fig. 3. It can be seen that a minimum potential between 0.55 and 1.0 volt must be applied to obtain linear photo-response. The maximum operating voltage that can be employed without visible hydrogen evolution at the cathode is about 4 volts.

The equilibrium values of the white light sensitivity attained with cells operated at different voltages are shown in Table II for both gray and black cathodes (platinum anodes used in all cases). The data show that there is little gain in sensitivity with increasing voltage above 2 volts

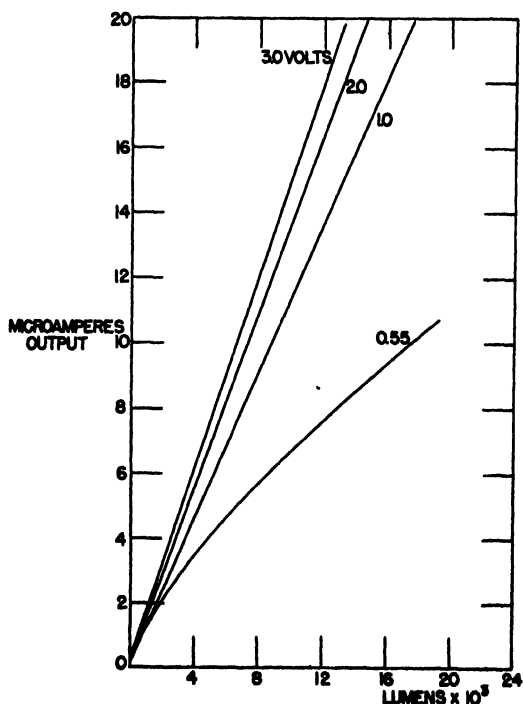


FIG. 3. Short-circuit sensitivity vs. operating voltage.

when black cathodes are employed. On the other hand, the dark current and the noise level (to be discussed below) increase rapidly with the applied voltage. Hence operation of the cell at an applied voltage of 2 v represents a good compromise. Unless otherwise stated, all data reported below refer to this operating voltage.

TEMPERATURE

A lower limit upon the temperature to which the photo-cell can be exposed is imposed by the freezing of the solution. As mentioned previously, this limit is -23° for a 57 weight percent solution of SeO_2 in water. No intrinsic upper temperature limit seems to exist below the boiling point of the solution, unless one is imposed by the particular cell housing employed. However, the dark current increases steeply with temperature (Fig. 4), while the light response, expressed as $I_{\text{light}} - I_{\text{dark}}$, remains constant within 10 percent over the

TABLE II.

Applied voltage	Sensitivity, $\mu\text{amp./lumen}$	
	Gray cathodes	Black cathodes
2.0	450 ± 100	1000 ± 100
3.0	800 ± 100	1100 ± 100
4.0	1250 ± 100	1350 ± 100

temperature interval given. On replotting the data of Fig. 4 in the form $\log I_d$ vs. $1/T^{\circ}\text{K}$, a straight line is obtained (Fig. 4a), the slope of which indicates an activation energy of about 0.38 electron volt for the conduction process. The variation of back e.m.f. with temperature may account for the slight deviation from the straight line at lower temperatures.

COMPENSATION OF DARK CURRENT AND NOISE LEVEL

In order to employ the photo-cell with feeble illumination where the corresponding photo-current might be considerably smaller than the dark current, it is necessary to compensate for the latter. This can be readily accomplished by employing the cell as one arm of a Wheatstone bridge as illustrated in Fig. 5. If R_1 is made high compared to the sum of the resistances of the measuring instrument, R_G , and R_s , and if the voltage across the cell is adjusted to be 2 volts, then the current output of the cell will be

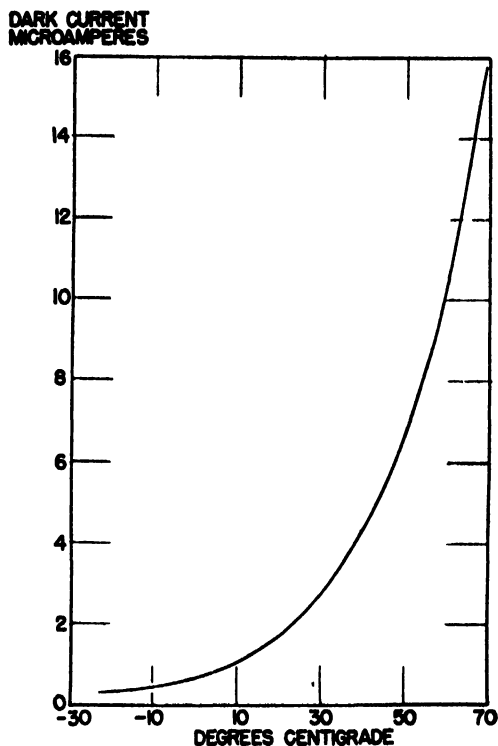


FIG. 4. Dark current vs. temperature.

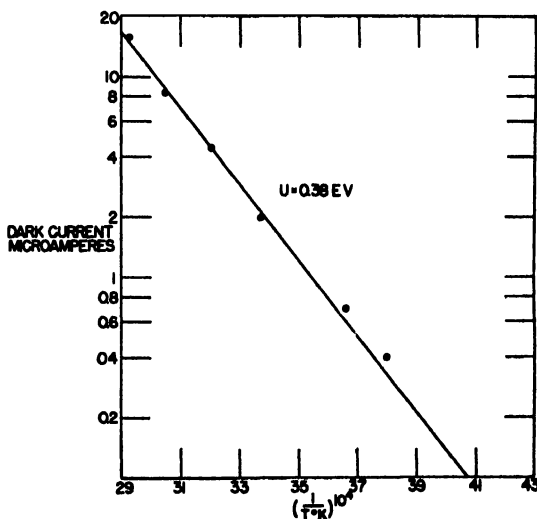


FIG. 4a. Temperature dependence of dark current.

identical with that shown in Fig. 7, where $R_{\text{external}} = R_G + R_s$. The bridge circuit might also be employed to compensate automatically for the temperature dependence of the dark current by replacing R_1 by another photo-cell having a temperature dependence of dark current identical with that of the original cell. Another possibility

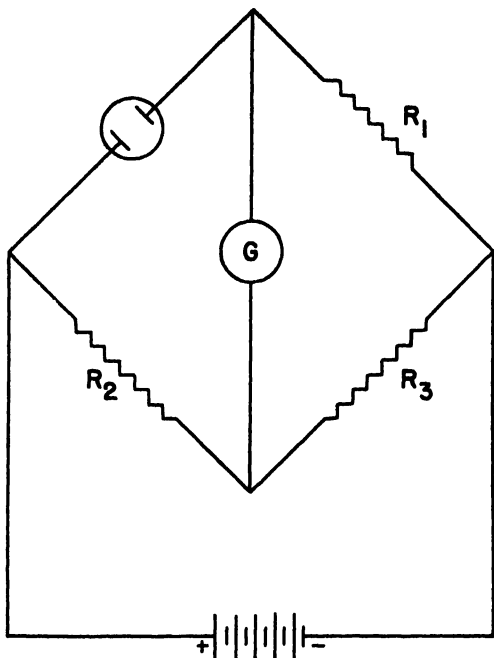


FIG. 5. Dark current compensation.

would be to employ a single cell with a common anode and two cathodes, replacing R_2 by the resistance between the anode and the added cathode. However, this scheme has the disadvantage that if R_3 is to be kept low, then R_1 must be made low in order to balance the bridge, resulting in the shunting of the meter and R_3 when off-balance current flows.

It was found that fluctuations of the order of 10^{-8} to 10^{-9} ampere exist in the dark current.

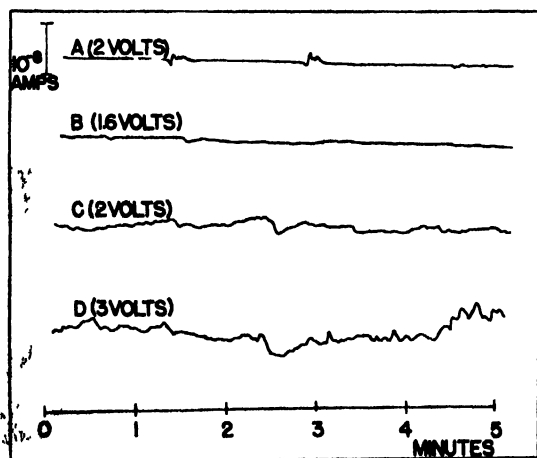


FIG. 6. Dark current fluctuations. A = cell No. 1. B, C, D = cell No. 2.

Figure 6 gives typical photographic records of the deflections of a sensitive galvanometer used as a null-instrument in the bridge with two different cells operated at different voltages. It is evident that some cells may have very few disturbances and others many, and that the frequency and magnitude of the "noise" increases with voltage, the signal equivalent of the noise being $1 \times 10^{-8} - 1 \times 10^{-6}$ lumen of unmodulated light.

LIGHT RESPONSE

Unmodulated Light

The current output of the photo-cell with a Se-plated anode as a function of unmodulated illumination with various values of resistance in

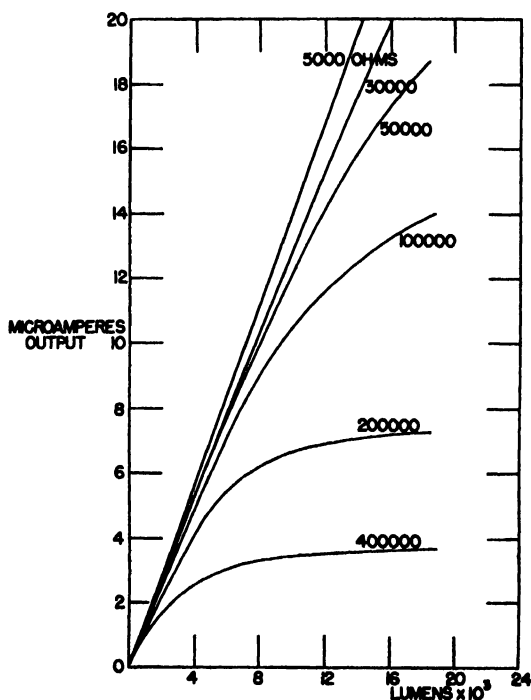


FIG. 7. Effect of illumination and external resistance on current output.

the external circuit and with 2 volts always maintained across the cell in the dark is shown in Fig. 7. The decrease in output and departure from linearity that occur with large values of resistance result from the fact that the cell acts as a photo-conductor with external driving voltage. With increasing current this driving potential is reduced by the voltage drop across the external resistance, thus decreasing the cell's

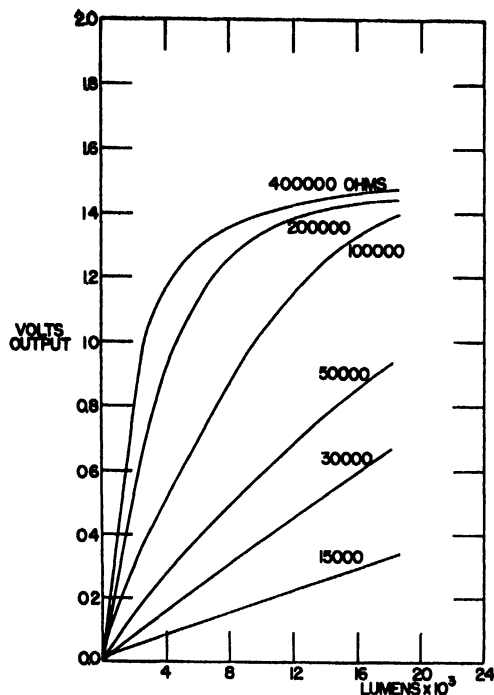


FIG. 8. Effect of illumination and external resistance on voltage output (2 volts applied).

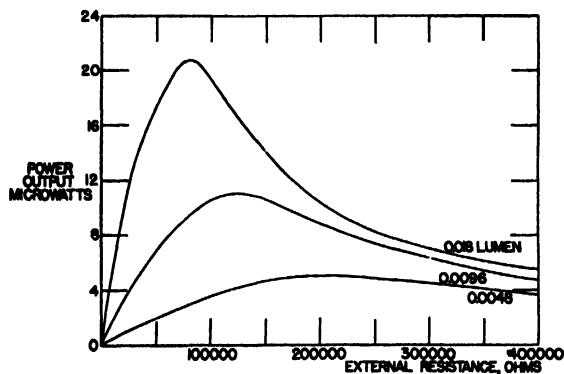


FIG. 9. Power output vs. external resistance and illumination.

sensitivity. Figure 8 represents the same data, recalculated in terms of the voltage output available across the external resistance, while Fig. 9 gives the power output as a function of external resistance for various values of illumination.

Marked deviation from linearity does not occur until the voltage across the cell falls below a minimum value of the order of 0.6 volt. Consequently, the higher the voltage applied across the cell plus load, the higher will be the voltage

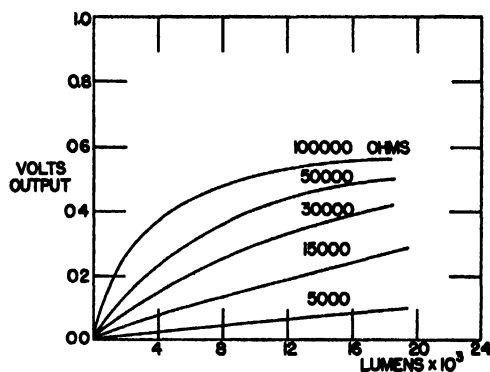


FIG. 10. Effect of illumination and external resistance on voltage output (1 volt applied).

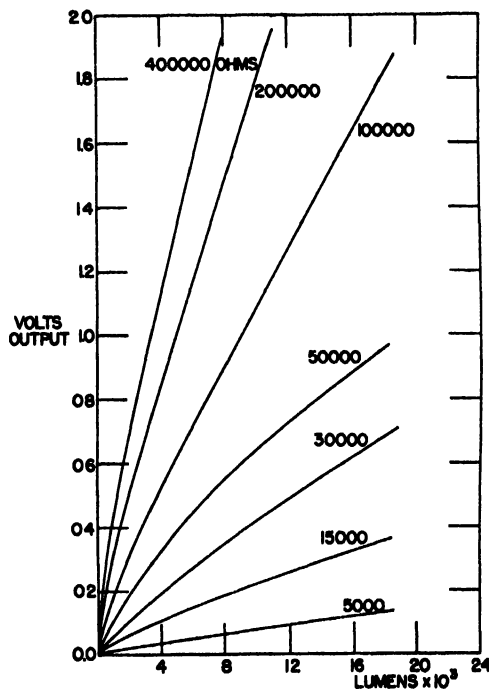


FIG. 11. Effect of illumination and external resistance on voltage output (3 volts applied).

available across the external resistance before attaining "saturation" (Figs. 8, 10, 11).

If the voltage across the cell is kept substantially constant by means of a feedback circuit such as that illustrated in Fig. 12, linearity of response results over a wide range of light intensities and external resistances (Figs. 13, 14). The elements included in the broken line in Fig. 12 are part of a bridge network that can be added to compensate for the dark current. The circuit, with minor modification, is generally

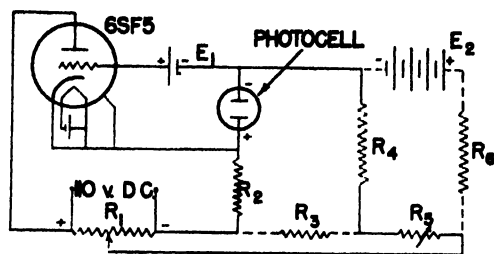


FIG. 12. Feedback circuit.

$E_1 = 1.5 \text{ v}$ $R_1 = 1500 \omega$ $R_2 = 10000 \omega$ $R_3 = 500 \omega$
 $E_2 = 22.5 \text{ v}$ $R_4 = 100000 \omega$ $R_5 = \text{Load resistance}$ $R_6 = 10 \text{ meg.}$

applicable to photo-conductive and photo-voltaic cells.⁵

Modulated Light

Measurements on the frequency response of the electrolytic photo-element were made using white light, sine-wave modulated, of 150μ lumens peak value. The circuit consisted of a cell, 2-volt battery and 30,000-ohm resistor in series, the resistance serving to feed the signal into an amplifier coupled to a General Radio Sound Analyzer. Typical data on cells that were several days old, but which had already reached equilibrium sensitivities to unmodulated light, are given in Table III.

A completely fresh cell compared at 80

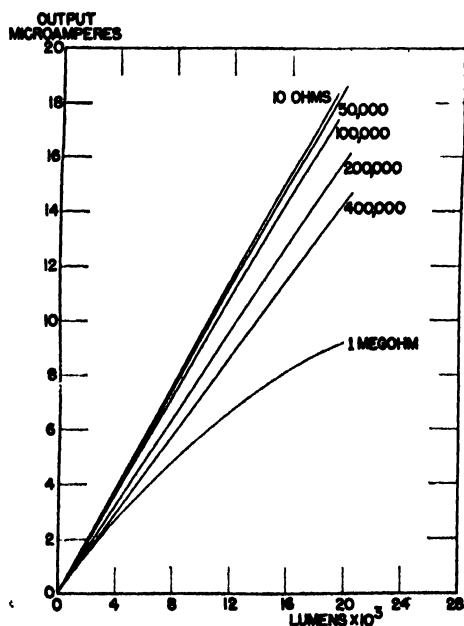


FIG. 13. Effect of illumination and external resistance on current output using feedback circuit.

⁵ A discussion will be given elsewhere.

cycles/second under similar conditions gives a signal output about three times as great and a noise level of about 1μ volt. An aging phenomenon occurs, however, as cells that have been stored on voltage for some months have a signal output only about one-tenth the value given in Table III at 80 cycles/second and a noise level of about 0.4μ volt, while the steady light sensitivity is not affected.

Additional data on several cells have been obtained by W. L. Hole at the University of Michigan with square wave modulation of 46μ lumens peak. For freshly prepared cells response to a 30 cycle/second signal was about one-sixth of that for the unmodulated signal (series resistance = 25,000 ohms), and the noise voltage amounted

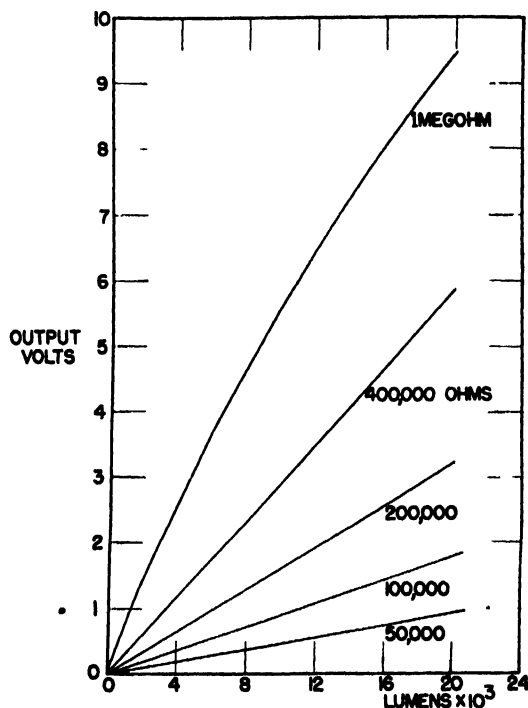


FIG. 14. Effect of illumination and external resistance on voltage output using feedback circuit.

to about 5μ volts. Relative response to different frequencies of signal modulation are shown in Table IV.

In addition it was ascertained that if the voltage across the cell is maintained at 2 volts, its response to weak, modulated signals becomes independent of strong unmodulated, background illumination. This can be automatically accomplished by the feedback circuit already described.

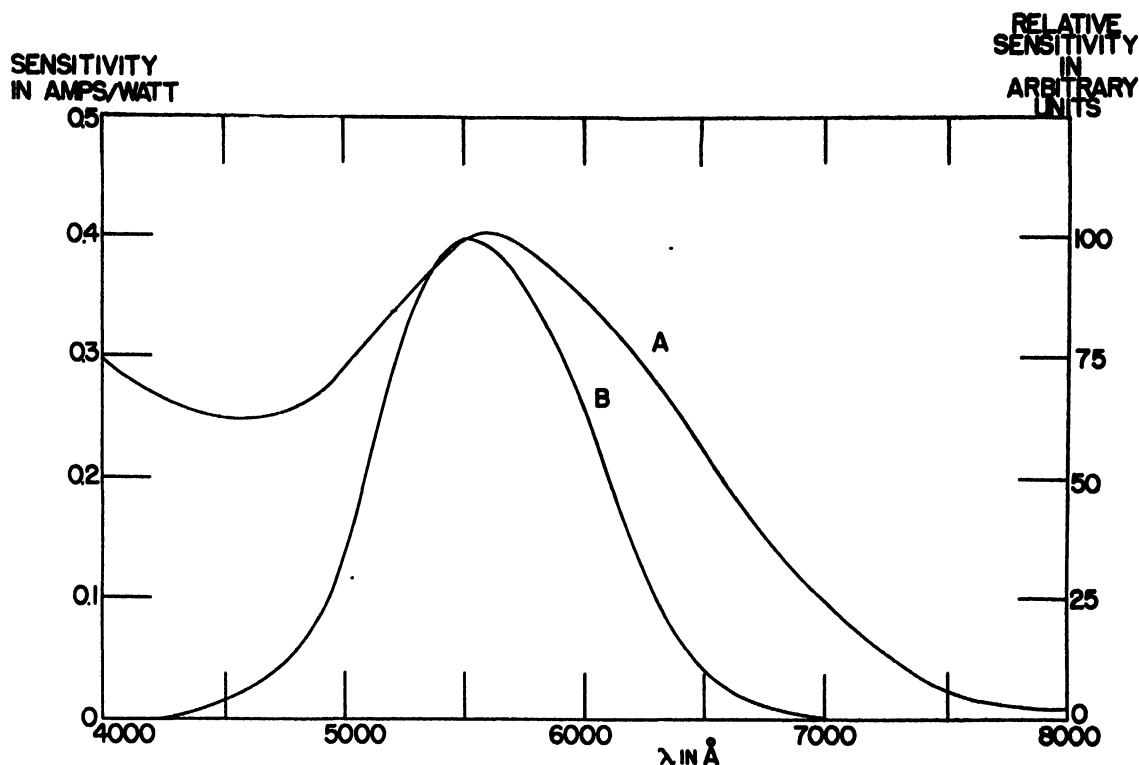


FIG. 15. Spectral sensitivity of *A*—Electrolytic selenium cell, *B*—Average eye.

Spectral Response

Constant energy spectral sensitivity data for the photo-cell (Fig. 15) were obtained by means of a double monochromator. An electronic regulator, which kept the light output obtained from the monochromator at constant energy by controlling the light input, permitted the automatic recording of the spectral response characteristic. It is of interest to note that the maximum in this characteristic corresponds to a quantum yield of about 0.9 electron per incident quantum.

LIFE STUDIES

Extensive studies on the life of the photoelement have been made with the following variations: platinum *vs.* selenium anodes; platinum, molybdenum, and stainless steel bases for the selenium; constant illumination and storage in the dark. At the time when the experiments were discontinued, cells containing platinum-base cathodes and either platinum or selenium on platinum anodes had maintained their original equilibrium sensitivity within about 15 percent for a period of six months. The same held true for

one such cell with a platinum anode that had been stored in the dark for three months and then constantly illuminated with a 150-watt lamp at 1-foot distance for three months. Cells with molybdenum-base cathodes and platinum anodes and stainless steel-base cathodes and anodes had maintained their high sensitivities during the test period of three months.

TABLE III. Frequency response of electrolytic selenium photo-element.

Frequency	Signal output	Noise level
sec. ⁻¹	μ volts	
0 (equivalent unmodulated signal)	2800	
80	36	~0.5 μ volt
200	19.5	
370	7.2	
535	5.8	

TABLE IV.

Frequency, sec. ⁻¹	30	90	330
Relative response, %	100	36	10

COMPARISON WITH OTHER WORK

A number of investigators⁶⁻⁹ have dealt with electrolytic selenium photo-cells and have observed on illumination of the sensitive selenium surface the development of an e.m.f. as well as the appearance of a short circuit current that quickly decayed. Reinganum¹⁰ has further observed that the cell's sensitivity could be increased and appreciable currents drawn from it if the selenium surface were made the cathode with respect to the other electrode by means of an external battery. An optimum operating voltage was observed using electrolytes of H_2SO_4 and HCl in agreement with our data of Table I. A cell almost identical with that of Reinganum was later described by Reiff.¹¹ Still more recently, a photo-voltaic cell, which we believe to be essentially similar to that of Reinganum, was described by Fink and Alpern.¹² This latter cell made use of a lead anode whose electrochemical counter potential with respect to the solution employed happened to be of the right sign and order of magnitude so that an external battery was not required. The current sensitivity of the cell was of the order of $150 \mu\text{amp./lumen}$.¹³

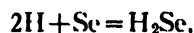
Our cell differs from the earlier electrolytic selenium cells chiefly in that directly electro-deposited metallic selenium gives rise to a higher sensitivity and that the selenious acid permits a higher lifetime as well as hermetical sealing of the cell when a selenium-coated anode is employed. Its characteristics rival most closely those of present commercial selenium "barrier layer" type cells. Its main advantages over the latter type are higher sensitivity, the possibility of making the active surface in practically any shape, and perhaps ease of manufacture. The need for an external voltage supply, the resulting dark current, and higher noise level, and the temperature

limit imposed by the freezing of the electrolyte are its main disadvantages.

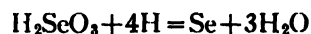
DISCUSSION

The resistance of a selenium electrode cannot be greater than the order of 100 ohms judging from the over-all conductance of the plating bath. However, the apparent resistance of the electrode in the photo-cell is of the order of a megohm, even if liberal allowance is made for the polarization e.m.f. of the element. The conclusion seems inescapable that a high resistant layer is formed at the cathode by the passage of current. Considering the electrolytes of Table I, the only likely electrode reaction at the cathode is the discharge of hydrogen ions. Hence, the layer may consist of an interstitial solid solution of hydrogen in selenium.

The gradual stripping of the cathode in cells containing acid electrolytes would result from the reaction



and similarly the plating of selenium at the cathode in selenious acid solution by hydrogen ion discharge would be the consequence of the reaction



or



The rapid decrease in dark current with time after the initial application of voltage is indicative of the building up of the high resistant layer. The large photo-sensitivity observed initially probably represents a lowering of the volume resistivity of the selenium deposit under the influence of light. After the layer of high resistance has been established, the volume resistivity of the selenium is negligible in comparison, and photo-currents observed at that stage probably represent a primary photo-effect.¹⁴

From a macroscopic viewpoint the presence of a thin layer of high resistance should give rise to an appreciable capacitance. The response of the photo-element to modulated signals can be explained on this basis.

¹⁴ The quantum yield for a cell operated at 2 v is quite close to unity at 4000 and 5600 Å and may slightly exceed 1 at higher voltages. However, primary photo-currents corresponding to yields of 2 electrons per quantum have been observed for the F' -bands in the alkali halides and a similar situation may prevail here.

⁶ K. Sabine, *Phil. Mag.* 5, 401 (1878); *Nature* 17, 512 (1878).

⁷ G. M. Minchin, *Phil. Mag.* 31, 207 (1891); *Lumière électrique* 48, 543 (1893); *Electrician* 26, 361 (1891).

⁸ C. Ries, *Das elektrische Verhalten des kristallinischen Selen gegen Wärme und Licht*, Dissertation, Munich (1902).

⁹ A. A. C. Swinton, *Phys. Soc. Proc.* 27, 186 (1915).

¹⁰ M. Reinganum, *Deutsch. Phys. Ges. Ber.* 4, 590 (1906); *Physik. Zeits.* 7, 786 (1906); *Physik. Zeits.* 8, 293 (1907).

¹¹ H. J. Reiff, *Electrician* 60, 467 (1908).

¹² G. Fink and D. K. Alpern, *Trans. Am. Elec. Soc.* 62, 369 (1932).

¹³ Using an electroplated selenium cathode, we have obtained a current sensitivity in this type of cell as high as $600 \mu\text{amp./lumen}$.

Studies of Plastic Flow of Steel, Especially in Two-Dimensional Compression

P. W. BRIDGMAN

The Physics Laboratories, Harvard University, Cambridge, Massachusetts

(Received November 21, 1945)

The experimental methods are described by which rectangular blocks are subjected to two-dimensional compressional stresses and at the same time flow is maintained two-dimensional with deviations of less than one percent. Within this range of flow a rather wide range of stress conditions is possible, so that the equations of plastic flow can be examined over a correspondingly wide range of conditions. It is found that two-dimensional flow is much more sensitive than three-dimensional flow to shearing instabilities initiated at the edges where there are singularities in the mathematical solution. By proper lubrication it is, however, possible to produce approximately homogeneous deformations in two-dimensional compression at least up to shortenings to two-thirds the initial length. It is found that the transverse stress X_z required to maintain two-dimensional flow is consistently higher than the theoretical value $\frac{1}{2}Z_x$; for some materials the excess above the theoretical value may rise to as much as 20 percent. This excess tends to be less in the early stages of flow, and does not tend to increase beyond strains of 10 percent. The failure of X_z to be equal to $\frac{1}{2}Z_x$ indicates a failure of the fundamental condition of isotropy of flow; this failure

may be demonstrated in other ways. The time rate of primary flow is studied at various points on the strain-hardening curve. The rate of flow rises rapidly as the stress increases above the limiting strain-hardening curve. The rate of increase of the rate of flow for a given displacement from the limiting curve is much greater in the early stages of flow. It is this which makes possible the calculation of the rate of propagation of a plastic disturbance from the parameters of the static strain-hardening curve. Beyond the early stages of flow new time effects begin to appear. The flow loses its smoothness and becomes more and more inclined to be jerky. At high strains, flow may not start at once when the load is increased, but there may be an initiation period during which flow is built up. The application of the results to the generalized strain-hardening curve is discussed. In view of the failure of isotropy, strictly speaking, a generalized strain-hardening curve does not exist. Within the strain limits of this paper it is found that both the maximum shearing stress criterion and the "significant" stress-strain criterion apply with errors of the order of 10 percent, the maximum shearing stress criterion being on the whole perhaps somewhat better.

INTRODUCTION

THE stress conditions under which plastic flow has been studied up to the present have been usually restricted because of the experimental inconvenience of applying stresses which are arbitrarily variable in different directions. The most common method of realizing more than one arbitrary component of stress has been by the simultaneous application of longitudinal tension and internal hydrostatic pressure to thin tubes. By this method it is possible, within certain limits, to apply two arbitrary components of tensile stress. The plastic flow which can be produced in this way is limited to comparatively small values by the incidence of tensile fracture. Much larger strains could be realized without fracture under compressive stresses, but few, if any, such experiments appear to have been made, although it would seem that it would not be especially difficult to expose a tube to external pressure with simultaneous longitudinal compression or tension. It would, however, probably be necessary to use tubes of appreciable thickness in order to avoid buckling

under the external pressure, and in this case the strains would not be homogeneous and the interpretation of the results would be complicated.

In the following, experiments are described in which rectangular blocks are exposed to arbitrarily variable compressive forces on two of the three pairs of faces, the remaining pair of faces being free from force. The stresses and the strains are homogeneous, and the mathematical interpretation of the results is especially simple, the experimental test block playing the role of the mathematical element of volume of the fundamental equations. The compressive forces are applied by hard platens pressing against the faces. Obvious geometrical limitations arising from mutual interference of the platens impose rather definite restrictions on the strains. There are doubtless various ways of meeting the geometrical conditions. In the following the strain as well as the stress was kept two-dimensional, on the gross average. Since two-dimensional stress in general produces three-dimensional flow, restriction of the strain approximately to two dimensions might appear to

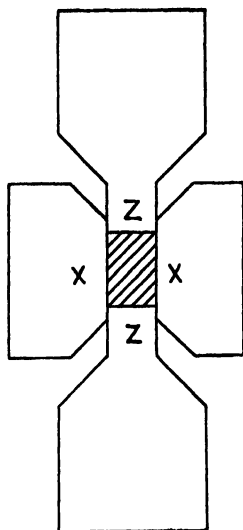


FIG. 1. Section through the platens for simultaneous compression in two directions at right angles. The specimen is shown shaded. The Z blocks and the X blocks are pushed together by two independent hydraulic presses.

seriously limit the part of the two-dimensional stress domain open to exploration. However, small deviations from a rigid two dimensionality in the strain were instrumentally feasible, which gave sufficient flexibility to permit a very considerable variation in the stresses, so that a very appreciable part of the total region of plastic flow under two arbitrary compressive stress components could be explored in spite of the geometrical restrictions.

THE APPARATUS AND METHOD

The direction of larger compressive stress will be denoted by z , that of minor compressive stress by x , and the direction of no stress by y . The corresponding stresses will be written as Z_x , X_z , and $Y_y (=0)$. The Z_x stress is applied by a pair of opposing rectangular blocks, of indefinite length in the y direction and with a width in the x direction approximately equal to the x dimension of the specimen. The X_z stress is applied by two hard blocks of indefinite extent in both the z and the y direction; the distance between these two blocks in the x direction must be kept not far from the initial width of the test block. The arrangement is indicated in Fig. 1. If the distance between the x blocks becomes less than the initial x dimension of the z blocks, these will be pinched and the measured stresses will be in

error. On the other hand, if the distance becomes too great, the test block will flow plastically into the narrow channel between z and x blocks, destroying the homogeneity of both stress and strain. In the following experiments the initial x dimension of the test block was made 0.502 inch and of the z block 0.500 inch. During the experiment the distance of separation of the x blocks was kept between 0.501 and 0.504 inch; within these limits there was neither pinching nor appreciable plastic flow into the crevasse at the edge, which was only 0.002 inch wide at the most. The x component of force must be maintained in such adjustment as to keep the x dimension in these limits; a rather wide latitude proved possible in meeting these conditions.

The z blocks and the x blocks were pushed together by two independent hydraulic presses. The press driving the z blocks was a standard press that has been used in much of my other work; it has a 3.5-inch plunger, a free space of 10 inches between the rods, and is driven by a hand pump with capacity of 15,000 p.s.i. The x blocks were pushed together with a specially constructed press with 2.5-inch plunger. This press was machined from a single piece of steel 4 inches in diameter and was compact enough to be mounted transversely in the 10-inch free space between the tie rods of the 3.5-inch press. Pressure was furnished by a second hand pump. This press was freely suspended from a point in the vertical line with its center of gravity, so that the whole press floated and could follow the motion of the specimen as it was plastically deformed, without exerting any extraneous force on it. The maximum displacement of the x press was 0.002 inch in the x direction and 0.125 inch in the z direction. Pressure was led into it through copper tubing of 0.25-inch outside diameter and 0.062-inch inside diameter; the tube was made long enough to exert no appreciable constraint.

Two different methods of applying the loads were employed. In the first, the pistons were advanced by pumping until approximately the desired increment of strain was reached. Pumping was then stopped and time allowed to elapse until a steady state was approximately attained. This steady state was reached by a simultaneous further slight automatic change of both strain

and pressure. Since the flow in the z direction was, in general, many-fold greater than that in the x direction, most of the manipulations were connected with the z press. A fraction of a stroke on the pump of the x press at the beginning of the stop usually was sufficient, the pressure of the x press then automatically taking care of itself. The steady-state pressures were read on calibrated Bourdon gauges in the lines of the two presses. In the second method of applying load, measurements were made at controlled intervals of total load. This was accomplished by interposing in the lines leading to the presses two dead weight free piston gauges. Pressure was adjusted by setting the appropriate weight on the pan of the dead weight gauge, and pumping until the piston floated. Since the motion in the x direction was so slight, automatic rise or fall of the free piston of the x gauge was sufficient to maintain the load constant within the limits of x deformation. The motion in the z direction was sufficiently large, however, to demand frequent pumping to keep the z free piston floating. For the more rapid rates of flow in the z direction the sluggishness of response of the free piston gauge was great enough so that the load would have fluctuated within undesirably wide limits even if the piston were kept always floating by rapid pumping. Especial measures had to be taken to meet this condition. In the first place, the piston of the free piston gauge was made much looser than is convenient for normal operation. The most important modification, however, was to provide volume flexibility in the line to the z press by attaching a reservoir of 750-cm³ capacity containing compressed nitrogen, obtained initially from a storage flask carrying 1500 p.s.i. Up to pressures on the 3.5-inch piston of 1500 p.s.i., pressure was obtained from the compressed storage flask by admitting the requisite amount of gas through a valve. Above 1500 p.s.i., the storage flask was cut off and pressure obtained by the hand pump. The maximum pressure on the 3.5-inch plunger was of the order of 7,000 p.s.i. A simple calculation showed that even under the most rapid plastic flow of these experiments inertia effects in the moving parts of the presses are inappreciable.

The running dimensions of the test block in the z and x directions during the course of the ex-

periment were measured by measuring the distance of separation of the z and the x blocks. This was done with Ames dial gauges, mounted on yokes attached to slender rods straddling the specimen and passing through appropriate holes drilled in the blocks. Since the motion in the z direction was large, reaching a maximum of 0.25 inch, it was sufficient to measure the z displacement with an ordinary gauge, graduated to 0.001 inch, with 1-inch total stroke. The x displacements, being much smaller, were measured with a jewelled gauge graduated to 0.0001 inch with total stroke of 0.2 inch. Any lagging of the gauge readings due to friction was eliminated by introducing into the whole system a very slight amount of mechanical vibration by pressing an electric razor against an appropriate member of the apparatus. Consistent differential readings of the x dimensions could be made to 0.00001 inch or even better if the pointer of the dial happened to be exactly on a division.

Under stress, the dimensions of the test block differ from the dimensions indicated by the Ames gauges by the elastic distortion of the z and the x blocks. A special study was made of the elastic distortion by stressing hardened test blocks of various dimensions. Any correction of this sort was found to be unimportant. The method finally adopted for dealing with it was to obtain the elastic distortion at the maximum stress by comparing the Ames gauge readings with the micrometer dimensions of the test block after removing from the apparatus, and to distribute the maximum correction over the intermediate stresses in proportion to the load.

The x blocks were made of tool steel, hardened to Rockwell C 67. In general the stress which the x blocks have to carry is one-half that of the z blocks. The z blocks were also at first made of the same tool steel as the x blocks. Their performance was later somewhat improved by setting in carboloy inserts. Carboloy can be highly polished so as to minimize the friction; furthermore, carboloy, being only one-third as deformable as steel, does not permit the test specimen to imbed itself to the same extent, so that the effective friction is reduced as the test specimen extends itself plastically along the platen, pushing its advancing edge up the incline of the depression in the platen. •

One of the objects of the experiments was to study the rates of flow for various combinations of stress. The "flow" understood here is the "primary" flow which occurs immediately after application of an increment of load and does not include such long range phenomena as elastic after-effects and creep. Of course, there is no sharp dividing line between these various effects; the time intervals involved in these measurements were not greater than the order of fifteen minutes. At first observations were made with manual recording and an ordinary watch, but these proved inadequate, and the final study of the rates of flow was made with photographic recording. The two Ames dial gauges were mounted side by side in the field of an Eastman Cine Special, for 16-mm film, so arranged that single frames could be exposed, either by hand operation for moderate rates of flow, or for more rapid rates by motor control, permitting up to a maximum of four exposures per second. The full opening, $f:1.9$, was used, as well as maximum film speed, corresponding to 64 frames per second for normal operation, and one-quarter shutter opening. Illumination was by two suitably mounted 500-watt floodlights. It was necessary to use a film of high contrast, in which the exposure latitude was accordingly small; the film used was Eastman positive. I am much indebted to Mr. Paul Donaldson for skillful development of the film and advice with regard to photographic details. The time was provided by a synchronous Telechron motor connected so as to give one revolution per second, the hand sweeping over a dial graduated to 1/100's. The total number of revolutions (i.e., seconds) was counted by a conventional Veeder counter mounted so that it was also in the field of the camera. The current weights on the free piston gauges were also recorded photographically, suitable labels being placed manually in the field when the weights were changed. An additional check on the z load was obtained by also recording the stretch of the tie rods of the z press; this stretch gives the actual load on the specimen without the component due to friction of the ram. The stretch of the tie rods was measured with an Ames 0.0001-jewelled gauge, connected to the tie rods through a lever giving fivefold multiplication, so that changes of length of 0.000002

inch could be established. This corresponds to a Z , stress on the test specimen of about 2 kg/cm². With the tie rod gauge it was possible to establish that there was perceptible sluggishness in the response of the stress to changes of load on the free piston gauge, in spite of the amelioration by the use of the gas reservoir. This sluggishness, however, was confined to the initial stages immediately after change of load when the rate of flow was high, and was therefore not a matter of importance. On the conclusion of a run, the specimen was removed and several photographs taken of it on the same film, as a matter of record. The film was then developed, transferred to a couple of spools which permitted it to be conveniently examined under a microscope, and the desired data read off from the film and recorded. The photographic record was sharp enough so that times could be read to a fraction of a hundredth of a second, and the Ames gauges to a tenth of their smallest divisions.

THEORETICAL BACKGROUND

The experimental procedure was guided by the accepted equations of plasticity; investigation of the validity of these equations was one of the principal objects of this work. We shall write these equations in the form for flow:

$$\frac{\dot{\epsilon}_x}{X_x - \frac{1}{2}(Y_y + Z_z)} = \frac{\dot{\epsilon}_y}{Y_y - \frac{1}{2}(Z_z + X_x)} = \frac{\dot{\epsilon}_z}{Z_z - \frac{1}{2}(X_x + Y_y)}.$$

In these equations the x , y , and z axes are assumed to be the principal axes both of stress and of strain. This demands that the axes do not rotate in space during the straining process; this condition is obviously satisfied in the present experiments. The strains are the "natural" strains, that is, $\epsilon_x = \log_e(l/l_0)_x$. The usual sign convention will be followed; a tensile stress will be taken as positive and a strain of extension positive. Under the conditions of our experiments, Y_y vanishes identically, so that the equations can be specialized as follows:

$$\frac{\dot{\epsilon}_x}{X_x - \frac{1}{2}Z_z} = \frac{\dot{\epsilon}_z}{Z_z - \frac{1}{2}X_x}.$$

Furthermore, Z_z is always larger numerically than X_x . If we put $X_x = \alpha Z_z$, α will vary between zero and unity. Under these conditions $\dot{\epsilon}_x$ always has the sign of Z_z . The flow along the x axis

vanishes for $\alpha = \frac{1}{2}$, has the same sign as $\dot{\epsilon}_z$ for α greater than 0.5, and the opposite sign if α is less than 0.5. The ratio of the two velocities expressed in terms of α is: $\dot{\epsilon}_z/\dot{\epsilon}_x = (2\alpha - 1)/(2 - \alpha)$.

In the following experiments the conditions are investigated under which there is no flow in the x direction, and also, when there is such flow, whether the velocity of flow in the x direction bears the theoretical ratio to velocity in the z direction in terms of α . These questions are to be investigated at various points on the strain hardening curve; it is conceivable that one relation might hold in the initial stages of flow and another in the later stages when there has been appreciable strain hardening. The basic physical postulate back of the above expressions for flow is the postulate of isotropy. One of the objects of the following experiments may therefore be phrased: to find to what extent the basic assumption of isotropy is satisfied.

It is obvious that if α varies arbitrarily during the straining process the flow equations cannot be simply integrated, and we may *not* write the conventional expression

$$\frac{\epsilon_x}{\epsilon_z} = \frac{X_z - \frac{1}{2}Z_z}{Z_z - \frac{1}{2}X_z}.$$

In these experiments α did vary during the straining process, so that the simple integrated form is not admissible. In general, it is obvious that no integrated form of the equations, in which the history of the variation of α or something equivalent to it does not appear, can be better than an approximation. The mere existence of an integrated form of the flow equations obviously imposes restrictions. It is customary to suppose that not only are the restrictions met, so that the history of α does not enter the final equations for strains, but that there is such a simple connection between strains and stresses of different types that if the relation between strain and stress is known for one type of stress it may be inferred for other types of stress. Various suggestions have been made as to the nature of the generalized relation. One of the most recent, and also one of apparently wide applicability, is that the "significant" stress, $\bar{S}$, is a universal function of "significant" strain, $\bar{\epsilon}$, for all types of plastic flow. The significant stress

and strain are defined as follows:

$$\begin{aligned}\bar{S} &= \left\{ \frac{1}{2} [(X_z - Y_z)^2 + (Y_z - Z_z)^2 + (Z_z - X_z)^2] \right\}^{\frac{1}{2}} \\ &= Z_z (1 - \alpha + \alpha^2)^{\frac{1}{2}}, \text{ (if } Y_z = 0, X_z = \alpha Z_z) \\ \bar{\epsilon} &= \left[\frac{2}{3} (\epsilon_x^2 + \epsilon_y^2 + \epsilon_z^2) \right]^{\frac{1}{2}} = \epsilon_z (4/3)^{\frac{1}{2}}, \\ &\quad \text{(if } \epsilon_x = 0, \epsilon_y = -\epsilon_z).\end{aligned}$$

The curve which gives $\bar{S}$ as a function of $\bar{\epsilon}$ may be described as the generalized strain-hardening curve. One of the objects of the following experiments was to determine the generalized strain-hardening curve under the present conditions (i.e., strain approximately two-dimensional, with two-dimensional stress), and to find whether significant stress is indeed a universal function of significant strain by comparing with the plastic flow obtained under other conditions of stress, in particular under simple compressive stress ($X_z = Y_z = 0$).

Returning now to the rate of flow, we have so far considered only relative rates of flow in different directions. There is also the more fundamental matter of absolute rate of flow. To consider this question the flow equation must be expanded into the form:

$$\frac{\dot{\epsilon}_x}{X_z - \frac{1}{2}Z_z} = \frac{\dot{\epsilon}_z}{Z_z - \frac{1}{2}X_z} = \varphi.$$

A complete investigation of the phenomena demands that we find of what arguments φ is a function, and what is the value of the function for various values of its arguments. It is obvious physically that for points in the stress-strain plane below the strain-hardening curve φ must vanish identically, there being no plastic flow here, and that φ must be everywhere positive in a region above the stress-strain curve. The mere existence of a strain-hardening curve would seem to suggest that φ gets larger very rapidly as the distance from the strain-hardening curve increases. Too great displacement from the strain-hardening curve would lead into the domain of fracture. φ must then, at the very least, be a function of stress and strain. It is such a function that for every value of stress, within limits, there is some strain which makes φ vanish; in general, the larger the stress, the larger the corresponding strain. In addition to being a function of stress and strain, φ may also well be a function of history. If it is not a function of history, as

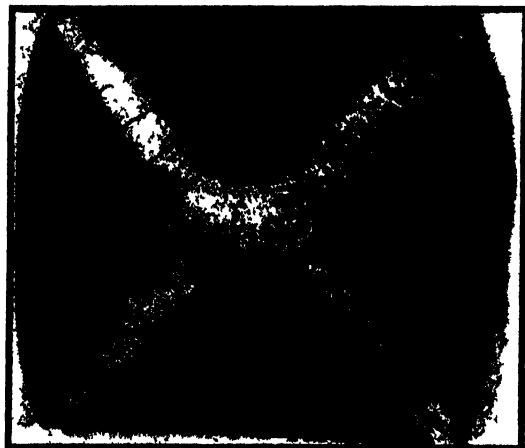


FIG. 2. Typical deformation with high friction.

perhaps we may assume provisionally as a first approximation, then $\varphi=0$ is the equation of the strain-hardening curve. Considerations of symmetry demand that under these conditions the stresses and strains enter through their invariants, or φ have the form $\varphi\{\text{stress invariant} - f(\text{strain invariant})\}$. Comparison with the definitions for significant stress and strain shows that this is equivalent to $\varphi\{\bar{S} - f(\bar{\epsilon})\}$.

It was one of the purposes of the following experiments to get some indications as to the form of φ and to find to what extent the above simple assumptions about its nature are justified.

In actually carrying out the measurements the conditions are by no means as simple as in the ideal mathematical discussion above. The chief disturbing factor is terminal friction on the z faces. It is well known that in a simple compression test, with free lateral expansion, the specimen has a tendency to barrel out because of the frictional constraint on the ends. If friction is reduced by proper lubrication, barrelling disappears and the strain becomes homogeneous. Under our conditions, compression with one component of lateral flow prevented, the effects of terminal friction proved to be much larger, and were also much more difficult to eliminate. The nature of the distortion can be made visible by scratching on the X surfaces of the virgin specimen a rectangular grid, together with a couple of diagonals to make the distortion at the corners more manifest. Figure 2 is a photograph of the x face of a compressed specimen in which

the lubrication of the z face was not adequate. Part of the original y face has rolled up onto the z face, the edge still being, however, a more or less sharp right angle. This would suggest that there must be some sort of mathematical singularity at the edge. In fact, a moment's consideration shows that in a compressed specimen flowing over the platen with friction there must be a mathematical singularity at the edge. For in the z face, because of friction, there is at the edge a non-vanishing tangential stress Y_z , whereas in the y face at the edge the tangential stress component Z_y vanishes identically merely because the surface is free. Here is a contradiction, because the conditions of mathematical equilibrium demand that everywhere $Y_z = Z_y$. An exact mathematical characterization of the details of the singularity demanded to avoid this contradiction would depend in any particular case on the elastic deformation of the platen, and would doubtless be difficult to obtain. The general nature of the singularity, however, is probably the same whatever the details, and may be obtained from any particular solution. There is one particular solution in the literature due to Nadai.¹ Another particular solution, involving only very simple mathematical functions, has been found by Westergaard, but has not yet been published. Both of these solutions agree that the singularity at the edge involves a shearing stress which is a maximum on the plane approximately bisecting the edge angle. In the present work there was frequently observed a region of maximum shearing distortion in a 45° zone; it is plainly evident in Fig. 2. Slip in this zone gets started with great ease and propagates itself so as to amount to an essential instability. This instability may be favored by the geometry of the specimen. This sort of slip starts most easily if the x face is a square (z and y dimension equal). If the initial dimensions of the block are in the proportion of 3:2, 45° slip is not inclined to start at first, but as the distortion proceeds and the ratio of the sides passes through 1:1, diagonal slip is very likely to start. In fact, if the ratio of the y dimension to the z dimension is integral, a shearing pattern on 45° lines is very apt to develop; Fig. 3 suggests such

¹ A. Nadai, *Zeits. f. Physik* 30, 115 (1924).

patterns that have been observed for the 2:1 and the 3:1 ratio.

These regions of concentrated shear may be suppressed by suitable lubrication; the requirements are much more difficult to meet than they are for ordinary simple compression. This seems natural, for we are here concerned with slip on a single set of planes, whereas in simple compression, cross slip on two interfering systems of planes is involved, with greater destruction of the microscopic structure. In the description of the detailed experiments the details of the lubrication will be described, and also the investigations to find what error is introduced by the inhomogeneities that actually occur.

THE MEASUREMENTS

Three types of measurements were made: those in which an approximately steady state was reached, those in which the velocity of flow was measured under constant load, and those in which the rate of flow was maintained constant and the stress determined as a function of strain.

1. Approximately Steady-State Measurements

These measurements embrace a determination of the strain-hardening curve, and a determination of the ratio of X_z to Z_z (that is, of α) for vanishing flow in the x direction. This type of measurement was made first, and involved considerable preliminary work, the nature of which will be indicated.

The original program called for a series of measurements corresponding to those already made in simple compression; in these experiments the height of the block was reduced to $\frac{2}{3}$ its initial height in a single stage of compression, the block was then remachined back to its original proportions, and the process repeated seven or eight times until the total reduction was some twenty-fold. It was soon obvious that there would be great difficulties in carrying through an analogous program for two-dimensional compression because of the much greater tendency to distortion during compression. Even a single stage of compression with a reduction of height to $\frac{2}{3}$ demanded much more perfect lubrication. A number of experiments were made to find the best lubrication. The method used successfully

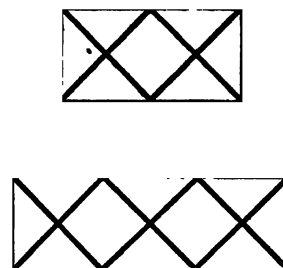


FIG. 3. Schematic diagram showing the regions of maximum shearing distortion of blocks compressed with friction on the bearing faces, with an integral ratio of height to breadth.

for simple compression was the same as that used for the lubrication of the external conical surfaces of my high pressure containers, namely, two thicknesses of 0.002-inch lead foil smeared on all sides with a thin paste of colloidal graphite, glycerine, and water. The distortion with this lubrication proved undesirably large under present conditions. Other simple lubricants were tried: Scotch tape, stearic acid, lead oleate, gold leaf, either alone or in various combinations with the lead foil, but were not satisfactory. A thin layer of soft solder applied with a soldering iron was not bad, but was not used for fear of annealing effects from the heat of the iron. Experience with the coat of solder was useful, however, in indicating that a plastic film tightly attached to the steel is desirable. If the film has not some sort of attachment it will be entirely squeezed out; this was the trouble with the lead. Various schemes were tried on the idea of a guard ring, that is, a separate layer of the same material in contact with the platens, the idea being that the deformation due to friction would be confined to the guard layer, leaving the central portion homogeneously strained. Any arrangement analogous to a guard ring proved, however, extraordinarily sensitive; it would either flow completely away from the central block or else embed itself in it, and the scheme was abandoned. The method finally adopted was to copperplate the z faces with a film of copper about 0.001 inch thick, and then put on top of the copper the sandwich of lead foil and graphite paste described above. Some difficulty was found in making the copperplating sufficiently adherent; the final technique was to start the plating in a cyanide bath and finish in an acid sulphate bath. The x

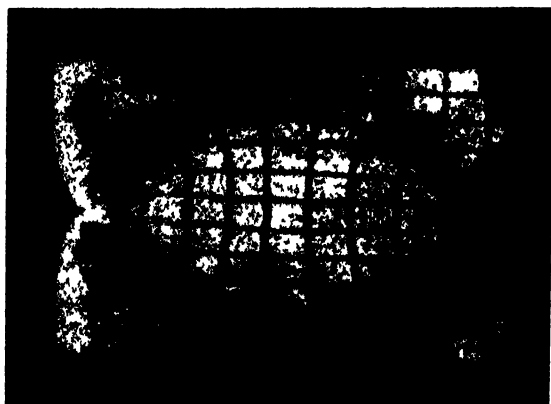


FIG. 4. Typical distortion of a block compressed for the second time after refiguring.

faces are not as sensitive to friction effects as are the z faces; for the x faces the simple lead foil sandwich was adequate.

After the first compression, the equator was always bulged slightly, even with the best lubrication. Under the best conditions the strain at the equator might be two or three percent greater than at the ends. If now the compressed specimen is machined back to the original proportions and subjected to a second compression, the equator at the beginning of the second compression will be more strain hardened, so that the ends will start to flow first. It might be thought that this condition would be self-correcting, since an excess flow at the ends would harden them and the flow would be thrown back to the equator. Actually, however, initial excess flow at the ends is followed by the development of shearing instability and the block folds into a crease based on the equator, as shown by Fig. 4. It might appear that this sort of thing could be avoided by making the end friction somewhat greater on the second application, but this proved too sensitive to manage, the second compression being almost certain to develop an instability like that of either Fig. 2 or Fig. 4, if the initial stage of compression was pushed as far as a shortening to $\frac{2}{3}$. It therefore proved infeasible to carry through the original program of pushing the strains to large values by the device of multiple compressions with refigurings. If, however, the strain of the individual stages is kept to something of the order of 10 percent, multiple compressions with refiguring can be carried out, and

runs were made on two different steels with six stages of compression with a total shortening to one-half.

The extra labor of making six stages in order to reach a strain of 0.7 (shortening to one-half) is not justified, however, because in a single stage strains up to 0.4 (shortening to two-thirds) can be reached with no appreciable error from lack of homogeneity of flow, although the inhomogeneity may be too great to permit a second stage of compression. The question of the character of the inhomogeneities and the error that might be introduced by them was examined with some care. The nature of the plastic flow was studied with the help of a network of squares scratched on the x faces, as already indicated. After compression the squares are deformed, and a measurement of the deformed squares under a micrometer microscope gives the local deformations. If the end friction is too great, the sort of plastic flow indicated in Fig. 2 is produced. The y face rolls up onto the z face at the edge, the angle at the new edge remaining comparatively sharp. The edges are, therefore, a region of infinite strain. From the edges a shearing cross is developed diagonally across the x faces; this shearing cross is marked by surface roughening of the steel, plainly visible in the photograph, and by great angular distortion of the squares. Based on the z faces are two triangular prismatic wedges which are pushed almost bodily into the remaining material. In this wedge the squares preserve nearly their original dimensions. At the center of the x face, where the shearing diagonals cross, the square is deformed to a rectangle, without angular rotation.

If the deformation were homogeneous, the initial height being h_0 and the final height h , it is easy to see that a square would be deformed to a rectangle, the ratio of whose sides is $(h/h_0)^2$. (For the height of the square is reduced in the ratio h/h_0 , and the width is increased in the ratio h_0/h .) The deviation of the ratio of the sides of the deformed square from $(h/h_0)^2$ may be taken as a measure of the failure of the deformation to be homogeneous. With good lubrication this ratio was found to have its theoretical value within experimental error. Thus, a block of soft Cr V steel compressed to

approximately two-thirds its initial height gave a ratio for the sides of the deformed square at the center of the x face of 2.23 against an $(h_0/h)^2$ of 2.28, and another block of the same steel under approximately the same compression gave 2.32 and 2.33, respectively. On the other hand a specimen in which there was a well-developed shearing cross and marked rolling at the edges gave for the ratio at the center of the x face 5.01 and at the base of the prismatic wedge 1.70, against an $(h_0/h)^2$ of 2.19.

Another method of measuring the failure of homogeneity, simpler and not so detailed, but adequate for most purposes, is merely to determine the ratio of the y dimension at the equator of the x face after compression to the average of the y dimension at the two ends of the x face. For good specimens this differs from unity by only a few percent.

In determining the strain-hardening curves it is the average stresses and the average strains which are determined, and it might be suspected that these averages would be relatively insensitive to local deviations from homogeneity. It turned out as a matter of fact that the averages were so insensitive that one could practically disregard any inhomogeneities within the limits of compression mentioned. Thus, in one experiment with a soft Cr V steel, the lubrication on the z face was intentionally omitted so as to develop the maximum inhomogeneity. The compressed specimen had a strongly developed shearing cross, and the rolling at the edges was so great that 16 percent of the final z face had originally been part of the y face. The strain-hardening curve for this specimen lay only 5 percent below the average of the curves for several other specimens, the deformation of which was essentially homogeneous. It was a general result that, if anything, the average stress for non-homogeneous distortion is less than for homogeneous distortion.

The general method of determining the strain-hardening curve with approximately zero velocity of flow in the x direction was as follows. The hand pumps were direct connected to the presses, without the intermediary of a free piston gauge, and the pressures were read from the corresponding Bourdon gauges. The dead spaces in the presses and connections were minimized so

that a maximum pressure increment would be produced by a given increment of strain. A preliminary run was made to determine the approximate values of the x pressure required for zero x flow corresponding to various z pressures and strains. Two observers took part in the final run; the function of one was to observe the x pressure gauge and of the other to perform the pumping and observe the z gauge. Both presses were first brought to pressures slightly below the initial plastic yield point. The x press was then raised somewhat, and the z pump then operated slowly. On passing the yield point the x pressure started to drop slowly, since it had been intentionally set too high, at the same time the x dimension decreasing. As the z pressure continued to rise with continued flow in the z direction, the drop of x pressure slowed down, ceased, and then reversed. Constant x pressure means zero velocity of x flow. The instant the x observer detected a reversal, the z pumping was stopped and the various readings recorded, thus obtaining the data necessary for a calculation of a point on the strain-hardening curve and for the simultaneous values of X_z and Z_z which make $\dot{\epsilon}_z$ vanish. The x pressure was then raised so as to be slightly "too high," and the process repeated. After familiarity had been acquired, it was possible by cautious procedure to obtain from 5 to 10 points on the strain-hardening curve in a total range of z strain of 0.4. This "familiarity" involved so choosing the increments of x pressure that the total flow in the x direction never passed the limits of 0.002 inch. No rigid time schedule was adhered to; the pumping was "slow," which meant that from 5 to 10 minutes was used in giving an increment of strain of 0.05. The x Bourdon pressure gauge was graduated to 10 kg/cm²; it was constructed by the Société Genvoise in the early 1900's, and had a very good multiplying mechanism, which gave readings with no perceptible back lash. It was possible to establish the point of reversal within 1 kg/cm² or better. From the known dimensions of the dead space in the x press and the compressibility of the transmitting oil it is possible to calculate that a pressure increment of 1 kg/cm² should correspond to a distortion of the specimen of approximately 0.000003 inch. That is, the pressure gauge should provide a sensitivity

in the detection of changes of the x dimension several times as great as that of the Ames 0.0001-inch gauge; this indeed proved to be the case.

The effects of friction were very small. The pistons of the presses were packed with the minimum amount of packing, and special tests showed that friction was not more than 2 or 3 percent under unfavorable conditions. Under the conditions of the experiment, the stress Z_z is smaller than that calculated with no allowance for friction because the z piston is advancing, overcoming both friction and the stress Z_z , whereas the stress X_z is greater than that calculated with no allowance for friction, because the stress X_z is advancing the x piston against friction and against the pressure of the x gauge. The calculations in the following were made with no allowance for friction. The corrected Z_z and X_z would be brought nearer together by a correction for friction; the maximum such effect is somewhere of the order of 5 percent.

The strain-hardening curves under conditions of zero x flow, with the simultaneous values of X_z were determined for the following materials: annealed "Solar" tool steel (typical composition: C 0.50, Mn 0.40, Si 1.00, Mo 0.50), eight different blocks cut in various orientations from a bar of 2-inch round stock; annealed Cr V steel (typical composition: C 0.40, Mn 0.25, Si 0.20, V 0.22), one series on 6 blocks of various orientations and with various methods of lubrication, one series of six compressings with refiguring on the same block, and another series on another block for four compressings with refigurings; a stainless steel of unknown composition, six compressions with refiguring on a single block; a 1045 steel provided by the Watertown Arsenal in the following heat treatments: austempered at 650°, 600°, 550°, 500°, and 450°, quenched and tempered at 650°, 550°, and 450°. Runs were made on one specimen of each of these heat treatments, or in a couple of cases on two or three. The initial dimensions of the specimens were: z , 0.750"; y , 0.500"; x , 0.502". The compressions, except for the multiple compressions with refiguring, were pushed to a reduction of height of two-thirds.

In addition to the data for two-dimensional flow, the strain-hardening curves in simple com-

pression ($X_z = Y_y = 0$, $\epsilon_z = \epsilon_y = -\frac{1}{2}\epsilon_x$) were also determined, either by special experiment or else taken from previous data.

The results are shown graphically in Figs. 5-13.

An examination of the curves shows that universally X_z is greater by up to 20 percent than its theoretical value, $\frac{1}{2}Z_z$. The discrepancy is greatest for the specimens austempered to the highest temperatures. Correction for friction would accentuate the discrepancy. In the initial stages of plastic flow X_z tends to be equal to $\frac{1}{2}Z_z$; the discrepancy reaches its maximum in the neighborhood of a strain of 0.1, and beyond this there does not seem to be usually any marked tendency for the discrepancy to become accentuated.

The result signifies, among other things, that the material is no longer isotropic after plastic flow in compression has taken place. In order to obtain some idea of the degree of failure of isotropy two other sorts of experiment were made. In the first of these, circular cylinders were cut from the compressed blocks with the axis of the cylinder along the original x direction, and then subjected to simple compression. If the material were isotropic, the section should remain circular on the second compression. Isotropy would hardly be expected, however, because one of the transverse dimensions is in the direction of previous maximum extension, whereas the transverse dimension at right angles to it is the direction of previous maximum shortening. Since the new lateral deformation is an extension, it would be expected that the strain hardening to the new deformation would be a maximum in the direction of previous extension. This would mean that the cross section of the circular cylinder would become elliptical, with the minor axis along the previous y direction. The experiment was tried with the stainless steel and annealed Cr V. With the latter, the distortion was in the expected direction, the minor axis of the ellipse being the former y direction. The maximum transverse strain in this particular experiment was 0.095 and the minimum 0.072. There was, therefore, a failure of isotropy by 32 percent ($0.095/0.072 = 1.32$), produced by an initial two-dimensional straining of 0.7 (the specimen was cut from the sample subjected to multiple com-

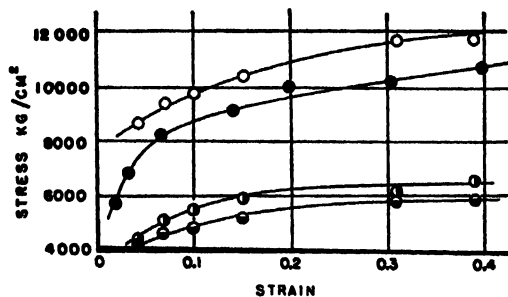


FIG. 5. Stress-strain relations for 1045 steel, austempered at 650°C. Reading from the top down, the first, third, and fourth curves relate to the block compressed, by suitable manipulation of X_z , so that flow in the x direction vanishes. The first curve plots Z_x , the third X_z , and the fourth $\frac{1}{2}Z_x$, which by elementary theory should be equal to X_z . The second curve is for another block with no lateral support.

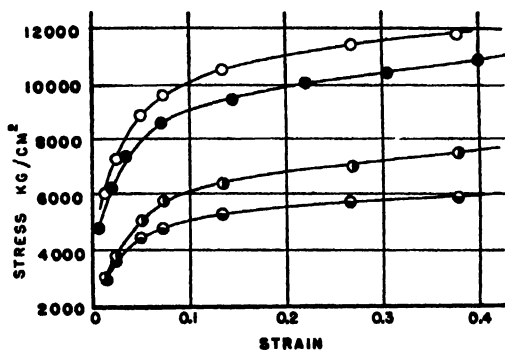


FIG. 6. Stress-strain relation for soft 1045 steel, austempered at 600°C. Reading from the top down, the first, third, and fourth curves relate to the block compressed, by suitable manipulation of X_z , so that flow in the x direction vanishes. The first curve plots Z_x , the third X_z , and the fourth $\frac{1}{2}Z_x$, which by elementary theory should be equal to X_z . The second curve is for another block with no lateral support.

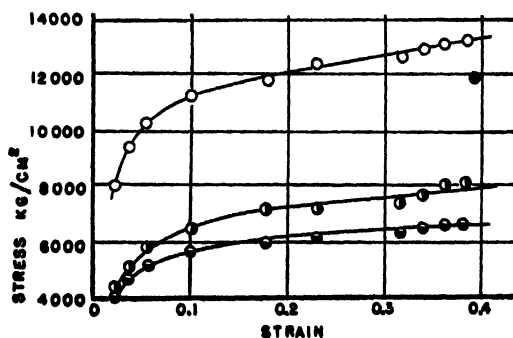


FIG. 7. Stress-strain relations for 1045 steel, austempered at 550°C. Reading from the top down, the first curve shows Z_x , the second X_z , and the third $\frac{1}{2}Z_x$, for a block compressed in such a way that flow in the x direction vanishes. This block was compressed in two stages; the points for strains greater than 0.2 are for the second stage, after the stress had been released, the block refigured, and again exposed to stress. The solid black disk shows a single point for another block compressed without lateral support.

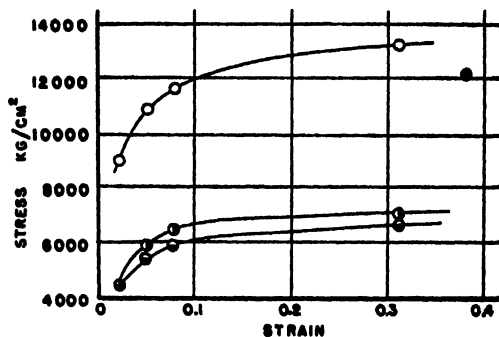


FIG. 8. Stress-strain relations for 1045 steel, austempered at 500°C. Reading from the top down, the first curve shows Z_x , the second X_z , and the third $\frac{1}{2}Z_x$, for a block compressed in such a way that the flow in the x direction vanishes. The solid black disk shows a single reading for another block compressed without lateral support.

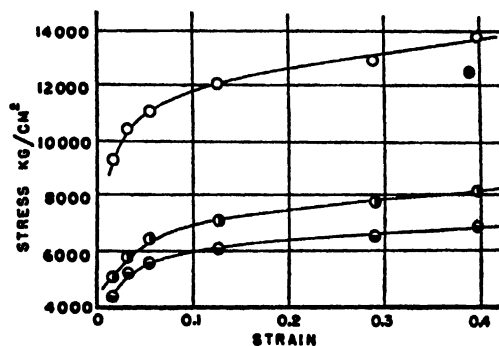


FIG. 9. Stress-strain relations for 1045 steel, austempered at 450°C. Reading from the top down, the first curve shows Z_x , the second X_z , and the third $\frac{1}{2}Z_x$, for a block compressed in such a way that flow in the x direction vanishes. The solid black disk shows a single reading for another block compressed without lateral support.

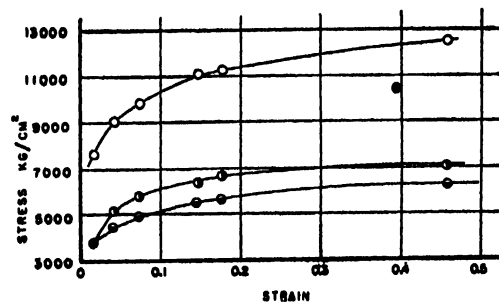


FIG. 10. Stress-strain relations for 1045 steel, quenched and drawn back to 650°C. Reading from the top down, the first curve shows Z_x , the second X_z , and the third $\frac{1}{2}Z_x$, for a block compressed in such a way that flow in the x direction vanishes. The solid black disk shows a single reading for another block compressed without lateral support.

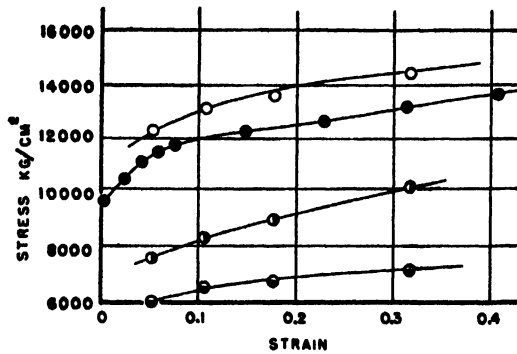


FIG. 11. Stress-strain relations for 1045 steel, quenched and drawn back to 550°C. Reading from the top down, the first, third, and fourth curves relate to the block compressed, with suitable manipulation of X_z , so that flow in the x direction vanishes. The first curve plots Z_x , the third X_z , and the fourth $\frac{1}{2}Z_z$. The second curve is for another block with no lateral support.

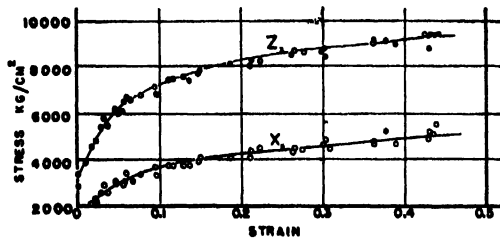


FIG. 12. Stress-strain relations for six different blocks of Cr V steel, X_z being so manipulated that flow in the x direction vanishes.

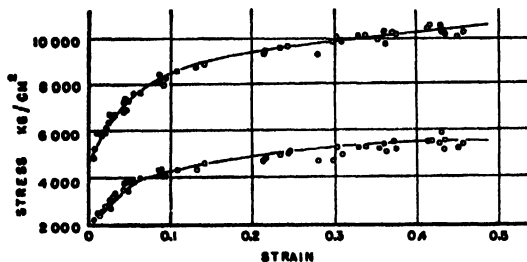


FIG. 13. Stress-strain relations for eight different blocks of "Solar" steel, X_z being so manipulated that flow in the x direction vanishes.

pression and refiguring). With the stainless steel, on the other hand, the expected relations were exactly reversed, the major axis of the ellipse being along the previous y direction. In this case the maximum and minimum transverse strains were 0.0675 and 0.0535, denoting a failure of isotropy in the unexpected direction of 26 percent ($0.0675/0.0535 = 1.26$). In other respects, also, the behavior of stainless steel in strain hardening is unlike that of the other steels investigated here; in particular, the relation between stress

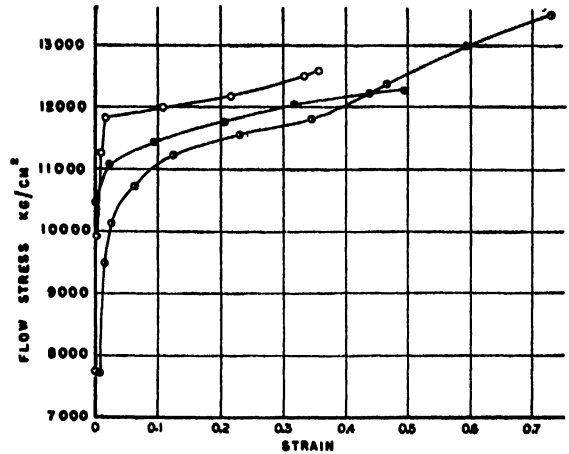


FIG. 14. Stress-strain relations in tension for small specimens of "Solar" steel cut in different orientations from a block previously strained in compression. Reading from the top down, the length of the first specimen is along the previous x direction (direction of previous zero strain), that of the second is along the previous y direction (direction of previous extension), and that of the third along the previous z direction (direction of previous compression). Fracture occurred at the final points of all curves.

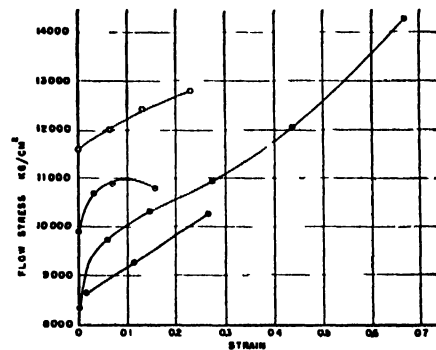


FIG. 15. Stress-strain relations in tension for small specimens of an annealed 1035 steel. The lower curve is for a virgin specimen. The three other curves are for specimens cut in various orientations from a block previously strained in compression. The top curve is for the specimen cut with its length in the previous y direction (direction of previous extension strain), the second curve for the previous x direction (direction of previous zero strain), and the third for the previous z direction (direction of previous compressive strain). Fracture occurred at the final points of all curves.

and strain shows much less curvature than for the other steels.

In the second test for failure of isotropy, small tension samples were cut from the compressed blocks in three different orientations, and stretched to fracture. The results are shown in Figs. 14 and 15. Again there is very marked failure of isotropy. The strain hardening for tension is least in the z direction, that is, in the

direction of previous compression. Furthermore, the strain to fracture in tension is greatest in the z direction. Both of these features seem natural, and suggest that there is a small reversible component in plastic flow. The strain hardening in the other two directions is relatively different for the two steels; the full situation is obviously complex.

In addition to the experiments with steel, a single run was made with a block of commercial annealed copper. In the initial stages of plastic flow simultaneous values of Z_x and X_z were 2,780 and 1,520 kg/cm², respectively; at the end of the test when ϵ_z had reached a value of 0.404, the corresponding values were 3,860 and 2,110. Again, as with steel X_z is somewhat more than one-half Z_x , indicating a failure of the perfect isotropy assumed in the equations. A marked difference between the behavior of steel and copper was that the steady state was reached more quickly with copper after pumping stopped.

2. The Velocity of Flow

The use of photographic recording in measuring the velocity of flow has already been described. The velocity measurements were made at constant load; this demanded the use of the free piston gauges, as already indicated. With constant pressure on the x press, the stress X_z is rigorously constant, because the area of the x face does not change during flow. The z face does, however, increase in area as flow progresses, so that constant z load does not ensure constant Z_x . The dependence of Z_x on ϵ_z at constant z load may be calculated as follows. We have:

$$Z_x = \text{load/section, section} \times z = \text{vol.}$$

Hence

$$Z_x = (\text{load/vol.})z,$$

$$(dZ_x/dz)_{\text{load}} = \text{load/vol.} = Z_x/z,$$

$$\epsilon_z = \log (z/z_0),$$

$$dz = z d\epsilon_z,$$

and

$$(dZ_x/d\epsilon_z)_{\text{load}} = Z_x.$$

The usual procedure was to keep the x load constant at some selected value, and determine the flow for successive increments in the z load.

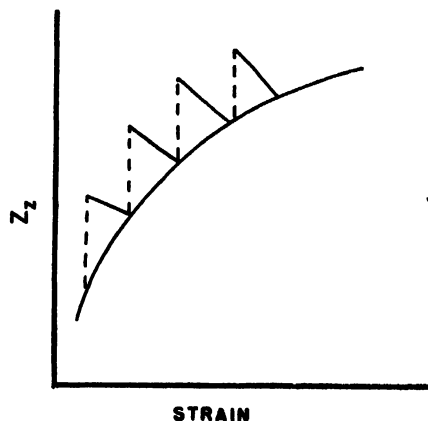


FIG. 16. Shows the general connection between stress and strain in determinations of rate of flow off the strain-hardening curve.

Each increment of z load was suddenly applied, by suddenly opening a valve to the gas reservoir in which the requisite pressure had been previously established. The initial stages of flow were comparatively rapid, and the camera exposures were made rapidly, at intervals of a second or less. The first exposure was made a fraction of a second after the application of load, as soon as it was safe to assume that inertia effects had vanished. As the flow slowed down, exposures were made at roughly equal increments of strain. Exposures were continued until the rate of flow in the z direction had dropped to something of the order of 0.0005 inch per minute. This procedure gives the velocity of flow at different points on the inclined parts of the sawtooth pattern indicated in Fig. 16. The slopes of the inclined part are given by $dZ_x/d\epsilon_z = -Z_x$ (ϵ_z drawn positive in the diagram). Successive increments of z load were made in this way at constant x load until the total deformation in the x direction had reached its permissible limit; the x load was now increased sufficiently to reverse the direction of x flow, and the step-wise increase of the z load was resumed. This was continued until the total shortening was about 33 percent. The total number of z steps varied from 10 to 25, average somewhere around 15, with a total number of photographic exposures of several hundred. Not all the exposures were usually used in taking the record off the film.

The flow phenomena of the following materials were studied: an annealed 1035 steel, five

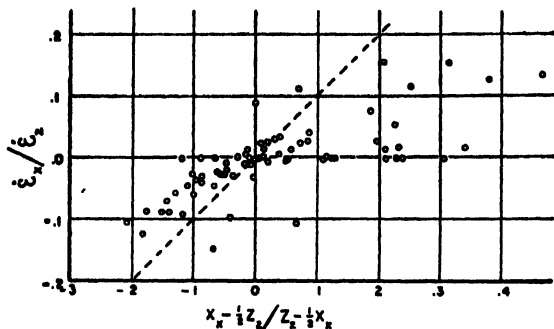


FIG. 17. Plot of all the observed ratios of rate of flow in the x direction to that in the z direction against the ratio of the "effective driving stresses" in these two directions. According to the elementary theory the points should lie on the dotted 45° line.

specimens in two-dimensional compression and one in simple compression; annealed "Omega" tool steel, typical composition C 0.55, Mn 0.80, Si 2.30, V 0.25, Mo 0.50, P 0.015, S. 0.025, four set-ups with two-dimensional compression and one with simple compression; stainless steel, one specimen in two-dimensional and one in simple compression; five heat treatments of the 1045 steel from the Watertown Arsenal, one specimen each in two-dimensional and in simple compression.

The photographic record gives the strains and the times. The rates of flow were calculated from the record by dividing increments of strain by increments of time. These rates were then plotted in one way or another depending on the point at issue. The photographic record contains the material for determining the strain-hardening curve from the limiting readings when the steady state was approximately reached. The strain-hardening curves were calculated from the photographic data and compared with those obtained by the first method of procedure. The agreement by the two methods was within experimental error; it is not worth while reproducing the details.

Although principal interest in the measurements was in the z rates of flow, the photographic record also contains the material for a calculation of the x rate of flow, but with less accuracy than for the z rate. This permits a calculation of $\dot{\epsilon}_x/\dot{\epsilon}_z$. Since the corresponding values of X_z and Z_z are known, we have the material for a comparison of the actual value of $\dot{\epsilon}_x/\dot{\epsilon}_z$ with its theoretical value in terms of the stresses, or in terms of α .

In Fig. 17, $\dot{\epsilon}_x/\dot{\epsilon}_z$ is plotted as a function of $R [\equiv (X_z - \frac{1}{2}Z_z)/(Z_z - \frac{1}{2}X_z)]$ for all the samples of steel for which $\dot{\epsilon}_z$ could be determined with any accuracy. The wide scatter of the points is understandable in view of the very small x strains. The conventional equations of plasticity demand that these points all lie on the 45° line. In spite of the scatter of the points, it is obvious that the 45° line definitely does not represent the results. So far as they can be represented by a line, the line has a considerably smaller slope, and furthermore passes somewhat below the origin instead of through it. This agrees with the results of the first part, namely $\dot{\epsilon}_x$ vanishes for a value of α somewhat greater than 0.5. No correlation could be found between the deviations of the points from a line and any other obvious single factor, such as grade of steel, location on the strain-hardening curve, or absolute value of the flow velocity. Only one point was found for which the *sign* of $\dot{\epsilon}_x/\dot{\epsilon}_z$ was definitely "wrong" in terms of the stresses. This particular point is hard to explain by experimental error, the increments of both ϵ_x and ϵ_z being too large to allow any chance for an error of sign. It may well be that the points actually should be scattered, and that the ratio $\dot{\epsilon}_x/\dot{\epsilon}_z$ depends on some combination of parameters more complicated than any that were obvious.

Much numerical material was collected for the velocity of primary flow, $\dot{\epsilon}_z$, and many curves were drawn. The maximum observed rate was 5×10^{-3} sec.⁻¹; the minimum was 0.00×10^{-3} . The material is too extensive and complicated to

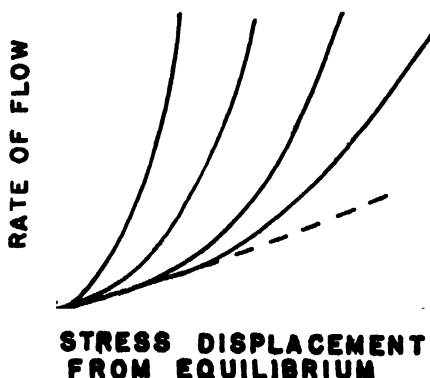


FIG. 18. Shows the progressive change in the character of the flow curves for increasing degrees of strain hardening.

attempt to reproduce in detail. There are, however, certain outstanding approximate generalizations. As seemed natural, the velocity of flow was found to increase rapidly with increasing distance from the limiting strain-hardening curve. To a rough first approximation the increase is exponential, that is, at points not too near the origin the logarithm of the rate is approximately linear in terms of the displacement from the strain-hardening curve. This displacement may be taken in terms either of stress or of strain. The rate at which the rate increases on moving away from the strain-hardening curve is itself a strong function of location on the strain-hardening curve, being much greater in the early stages of deformation, that is, for small strains. The qualitative nature of the effect is suggested in Fig. 18, the successive curves being drawn for the successive steps indicated in Fig. 16.

The abruptness of the rise of the rate in the early stages of plastic flow is suggested by the following example. The maximum strain rate measured in these experiments, $5 \times 10^{-3} \text{ sec.}^{-1}$, was for Omega steel at a total strain of 0.04, where the equilibrium stress is 8600 kg/cm^2 . This rate was produced by a stress displacement from the equilibrium curve of 130 kg/cm^2 or 1.5 percent. The rate just mentioned was the average over a time interval of 0.4 sec.; the initial rate must have been much higher.

The curves of Fig. 18 have a linear envelope, which approaches the origin with a finite slope. If the displacement from the strain-hardening curve is measured in terms of normal distance, instead of along the stress axis, the drawing apart of the curves becomes even more marked than in the figure. The general order of magnitude of the drawing apart of the rate curves may be suggested by a single example. For the 1045 steel, austempered at 650° , the rate of flow was $1.0 \times 10^{-3} \text{ sec.}^{-1}$ for a displacement from the strain-hardening curve of 13 kg/cm^2 in the early stages where the equilibrium stress was $5,800 \text{ kg/cm}^2$, whereas in the later stages, where the equilibrium stress was $11,000 \text{ kg/cm}^2$, a displacement from the strain-hardening curve of 470 kg/cm^2 , or 36 times as much, was necessary to reach the same rate of flow. The slope of the linear envelope shown in Fig. 18 depends markedly on the grade of steel, being less for the harder steels.

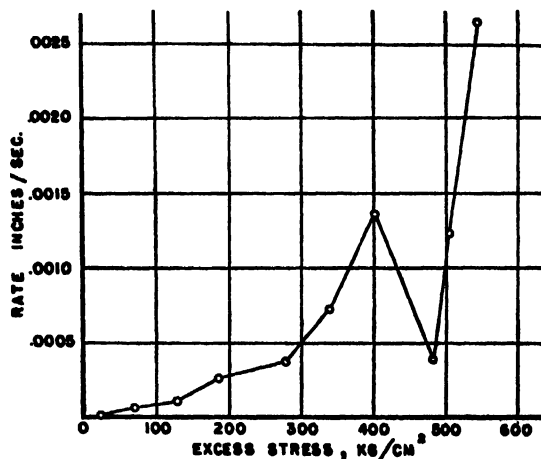


FIG. 19. Example, for soft "Omega" steel of regeneration of rate of flow, after drop from a high initial value.

As an example, this slope was $2.7 \times 10^{-6} \text{ cm}^2 \text{ sec.}^{-1} \text{ kg}^{-1}$ for the stainless steel, and $7 \times 10^{-7} \text{ cm}^2 \text{ sec.}^{-1} \text{ kg}^{-1}$ for the 1045 steel tempered at 450° .

No material difference could be found between the character of the rate curves for two-dimensional and simple compression. The very rapid increase in the rate of flow for slight displacements from the strain-hardening curve in the early stages of flow explains the possibility of approximately calculating the velocity of propagation of the plastic front from the statically determined stress-strain curve. The loss in the abruptness of the flow phenomena at higher strains, shown by the drawing apart of the curves in Fig. 18, would suggest that in a wire which had been previously severely strained the velocity of propagation could not be calculated from the stress-strain curve with as close a degree of approximation as for the virgin wire.

The actual rate curves, as a general rule, have the smooth idealized configuration shown in Fig. 18 over their entire extent only for the early stages of the strain-hardening curves, that is, only for small strains. For larger strains the smooth-rate curves are often interrupted by episodes with an abnormally high rate. This holds for both simple and two-dimensional compression. Figure 19 shows an example. The location of the episodes may be anywhere on the curve and would seem to be a probability affair. One is practically certain to find several of these episodes, if one waits long enough, in the region close to the strain-hardening curve

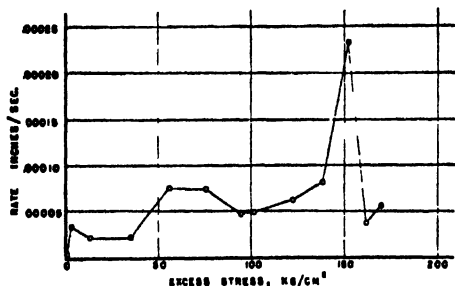


FIG. 20. Example, for soft "Omega" steel, of attainment of a relatively high rate of flow after a slow beginning.

where the rate on the average is approaching zero. It was not at all uncommon for the strain to hang at a steady value, constant within 0.00001 inch for one or two minutes, and then within 2 or 3 seconds increase by 0.001 inch.

This flow in jumps is not contained in the conventional equations of plasticity, but is obviously a matter of considerable importance in its suggestions as to the mechanism of plastic flow. There would seem to be little doubt that it is a real phenomenon and not an artifact. One might be inclined, at first to think it an effect of friction. But it cannot be friction of the x face, because the jumps are shown equally in simple compression where the x face is free. Neither can it be friction on the piston of the press, because the effect occurs only when the specimen has received comparatively large strains. Similar effects are found in other circumstances. One is reminded of deep-seated earthquakes. The continued internal fractures with self-healing which I have found when shearing is combined with high hydrostatic pressure are doubtless somewhat similar.³

It would be interesting and important to establish whether a jump in the z flow is accompanied by a jump in the x flow. The answer to this question could not be certainly found from the data, the absolute value of the x flow usually being too small. In two or three instances, however, there were slight jumps in the x flow accompanying those in the z flow. The only conclusion that is justified is that there is no present evidence against supposing that jumps in the two flows occur together, or in the theoretical ratio.

³ P. W. Bridgman, J. App. Phys. 8, 328 (1937); Proc. Am. Acad. 71, 387 (1937).

In addition to the flow in jumps there is another phenomenon sometimes found at the larger strains. On increasing the load suddenly at constant strain, flow is very likely not to begin at once, but the flow hangs and requires time to build up to its maximum rate. This is not an inertia effect, for it is found only in the later stages of flow, and the time intervals are too long. This effect was found only in the 1035 and Omega steels. An example is shown in Fig. 20. I have previously found a similar effect in tension. Tension specimens, pulled under hydrostatic pressure to an elongation far beyond the elongation at fracture under normal conditions, and then repulled at atmospheric pressure, may show a similar initial hanging of the flow. The mere existence of the effect shows that strictly the velocity of flow cannot be a function of stress and strain only, but must involve the history to some extent. The same conclusion, of course, could have been drawn from the existence of jumps in the flow.

For the more rapid rates of flow, the rate is not sensitive to X_z ; under the conditions of these experiments the regular sequence of curves indicated in Fig. 18 was usually not affected by a discontinuous change in X_z . At the smaller rates, where the effect to be expected from a change of X_z is more nearly of the same magnitude as the rate itself, the effect of changing X_z is more easily exhibited, and in particular the paradoxical effect already suggested, namely that an increase of X_z at constant Z_z may increase the rate of z flow. Figure 21 shows for the 1035 steel the effect of an increase of X_z in

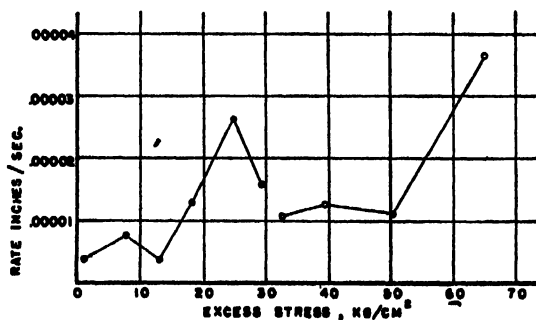


FIG. 21. Example, for 1035 steel, of the increase in flow velocity along z which may be brought about by an increase in X_z . At excess stresses greater than 30 the value of X_z was 4420 kg/cm²; below 30 it was increased to 5600. Z_z was approximately constant at 8100.

increasing ϵ_s . This increase of rate is contained in the equations of plasticity, as a consequence of an increase of the function φ due to an increase in the stress invariant part of the argument. Not all increases of X_s bring about an increase of φ , but obviously any increase of X_s from a value greater than $\frac{1}{2}Z_s$ to a value still greater than $\frac{1}{2}Z_s$ will result in such an increase; these are the conditions prevailing in Fig. 21.

3. Measurements with Constant Rate of Flow

Since the results of these measurements were mostly negative, they may be dismissed with a brief description. A constant rate of flow was maintained independent of the load by pumping liquid into the press at a constant rate. The primary source of motion was a "gerotor" gear pump, which is essentially an arrangement for injecting a constant amount of liquid per revolution. The useful range of the pump is only a few hundred pounds per square inch, at higher pressure the leak back through the gearing becoming serious. It was necessary, therefore, to interpose intermediate stages between the gerotor pump and the press deforming the specimen. The gerotor pump was arranged to drive an 8-inch piston, which drove a 1-inch piston, which in turn drove the 3.5-inch piston of the press. With this arrangement the pressures were kept within the necessary limits. The gerotor pump was driven through variable speed pulleys and gears by a motor of such large capacity that its speed was practically unaffected by variations of pressure in the range. The data were recorded photographically as before. The running load was taken from the stretch of the tie rods of the press.

Only simple compression was studied. Three different steels were examined, the 1045 steel tempered at 650°, 550°, and 450°. Each steel was examined at three rates of strain: 0.00212, 0.00080, and 0.000157 inches per second, that is, a range of 13.5-fold. Within this range the existence of an effect could be detected in the expected direction; the stress-strain curve for the highest speed lay higher than for the lowest speed by something of the order of 4 percent at the highest strains. The accuracy was not sufficient, however, to permit any more detailed

study. The sensitivity of the readings would have permitted this, but the results were made irregular by a comparatively large correction for the thermal expansion of the tie rods during the run by the heating of the flood lamps. It did not seem worth while to make the complicated modifications that would have been necessary to eliminate this source of disturbance, since the speed effect has been the subject of studies by other observers over a much greater range of speed and a wider range of conditions. One comment is perhaps worth making: It would be expected from the results in Part 2 that the effect of speed on the stress-strain curve would be more marked at large strains than at small ones.

THE GENERALIZED STRAIN-HARDENING CURVE

The problem of deducing the strain-hardening curve for one type of deformation from measurements on another type of deformation has been much discussed and is one on which the present measurements have some bearing. The two types of deformation with which we are concerned are simple compression and two-dimensional compression. Three possibilities in the way of generalizing the strain-hardening curve will be considered.

1. The maximum stress is a universal function of the maximum strain. The maximum stress is Z_s and the maximum strain ϵ_s for both simple and two-dimensional compression. If this function is the correct one, then Z_s plotted against ϵ_s should give coincident curves for simple and two-dimensional compression.

2. The maximum shearing stress is a universal function of the maximum shearing strain. For simple compression:

$$X_s = Y_s = 0, \quad \epsilon_s = \epsilon_y = -\frac{1}{2}\epsilon_x,$$

$$\left. \begin{aligned} \text{Maximum shearing stress} &= \frac{1}{2}Z_s \\ \text{Maximum shearing strain} &= \frac{1}{2}\epsilon_s \end{aligned} \right\}$$

For two-dimensional compression:

$$Y_s = 0; \quad |X_s| < |Z_s|, \quad X_s \text{ and } Z_s \text{ of same sign.} \\ \epsilon_z = 0; \quad \epsilon_y = -\epsilon_x.$$

$$\left. \begin{aligned} \text{Maximum shearing stress} &= \frac{1}{2}Z_s \\ \text{Maximum shearing strain} &= \epsilon_x \end{aligned} \right\}$$

If this is the correct function, Z_s plotted against $\frac{1}{2}\epsilon_s$ for simple compression should give the same curve as Z_s plotted against ϵ_s for two-dimensional compression.

3. The "significant stress," $\bar{S}$ is a universal function of the "significant strain," $\bar{\epsilon}$. The definitions are as follows:

$$\bar{S} = \left\{ \frac{1}{2} [(X_s - Y_s)^2 + (Y_s - Z_s)^2 + (Z_s - X_s)^2] \right\}^{\frac{1}{2}},$$

$$\bar{\epsilon} = \left[\frac{1}{2} (\epsilon_x^2 + \epsilon_y^2 + \epsilon_z^2) \right]^{\frac{1}{2}}.$$

For simple compression:

$$\bar{S} = Z_s, \quad \bar{\epsilon} = \epsilon_s.$$

For two-dimensional compression:

$$\bar{S} = Z_s(1 - \alpha + \alpha^2)^{\frac{1}{2}}, \quad \text{where } X_s = \alpha Z_s.$$

$$\bar{\epsilon} = \epsilon_s(4/3)^{\frac{1}{2}} = 1.155\epsilon_s.$$

If this is the correct function, Z_s plotted against ϵ_s for simple compression should give the same curve as $Z_s(1 - \alpha + \alpha^2)^{\frac{1}{2}}$ plotted against $1.155\epsilon_s$ for two-dimensional compression.

Under present experimental conditions there is one qualitative difference between the various criteria. If either of the first two is correct, Z_s plotted against ϵ_s for two-dimensional compression should give a curve passing smoothly without jump over any points where X_s may be changed discontinuously. If, however, the third criterion is correct, there should be a jump in the Z_s versus ϵ_s curves provided there is a jump in $\bar{S}$ on changing X_s , for $\bar{S}$ versus $\bar{\epsilon}$ is smooth. The larger the jump in $\bar{S}$, the more critical should be the test. Now $\bar{S}$ as a function of X_s or of α has a flat minimum in the neighborhood of $\alpha = \frac{1}{2}$, so that any jumps of X_s which keep it in the general neighborhood of $\frac{1}{2}$ will not afford a sensitive test. It is for this reason that all the quasi-static measurements described under the heading on page 231 are not well adapted to bring out the distinction that we are seeking. The curves specially determined for rate of flow by photographic recording in some cases embraced changes of X_s in a more favorable domain, and it is the limiting curves given by these measurements that were used in comparing the three criteria. It has already been mentioned that the limiting strain-hardening curves determined in this way were the same as those obtained by the first method.

As it turned out, the cases in which there were

favorably situated jumps in X_s did not permit a decision between the various criteria. In one instance, that of Omega tool steel, at a point where there was a jump in α from 0.45 to 0.72, there was a greater jump in the curve of $\bar{S}$ vs. $\bar{\epsilon}$ than in the curve of Z_s vs. ϵ_s , the former jumping up by 5 percent and the latter jumping in the same direction by only 2 percent. The 1045 steel, tempered at 650°, gave two opportunities for comparison. There was one jump of α from 0.38 to 0.68; at this point there was no discontinuity in the $\bar{S}$ curve, whereas Z_s jumped down by 3 percent. At a somewhat greater strain for the same material there was a jump down in α from 0.64 to 0.50. The $\bar{S}$ curve now jumped down by 2 percent, whereas the Z_s curve was sensibly continuous. Various points of discontinuity in α for other steels did not permit a decision between a jump or an abrupt change in direction of the curves, the experimental points not being sufficiently closely spaced. On the whole it appeared that the tendency to a break of some sort was greater on the $\bar{S}$ curves than on the Z_s curves.

The data would appear, therefore, not to permit a clean-cut decision on the basis of the breaks in the curves, and the decision must be sought from a comparison of the whole general course of the curves.

In making this comparison it is to be kept in mind that $\bar{S}$ and $\bar{\epsilon}$ are identically equal to Z_s and ϵ_s , respectively, for simple compression. In anticipation, neither of the criteria are met within the limits of error of the measurements, and it becomes a question of choosing the one which is the best approximation. As regards the criterion that Z_s is a universal function of ϵ_s , the Z_s curve for two-dimensional compression consistently lies above that for simple compression for all the steels except the 1035 steel. For strains above 0.1 the order of magnitude of the excess is 10 percent. For the 1035 steel the curves cross at a strain of 0.07, and at a strain of 0.15 the curve for simple compression has risen to 10 percent above the curve for two-dimensional compression. For all steels the curve of $\bar{S}$ for two-dimensional compression lies below that for simple compression. The discrepancy between the curves may be either greater or less than for the Z_s curves, but on the whole is inclined to be rather less, so that on the whole the $\bar{S}$ criterion is to be preferred to

the Z_s criterion. With regard to the maximum shearing stress criterion, the curve for maximum shearing stress for two-dimensional compression consistently lies above that for simple compression, except for the 1035 steel. The discrepancies are again of the general order of 10 percent, but detailed comparison would show that for the present selection of steels the maximum shearing stress criterion is on the whole the best by a narrow margin. For the stainless steel, either the significant stress or the maximum shearing stress criterion holds with much less error than for the other steels. In fact, the general behavior of the stainless steel is different from that of the others; its strain-harden-

ing curve, for example, is sensibly linear up to a strain of 0.5.

All of the above remarks apply in the region of strains of moderate magnitude, say above 0.02. In the initial stages of flow, with the single exception of the 1035 steel, the stress for two-dimensional flow is less than for the three-dimensional flow of simple compression. It is natural to associate this with an initial slip on a single set of slip planes in two-dimensional flow as contrasted with cross slip on systems of interfering planes in simple compression.

I am much indebted to Mr. Charles Chase for the skillful construction of the apparatus and assistance with the manipulations.

A Method of Determining the Size of Droplets Dispersed in a Gas

R. L. STOKER

Western Precipitation Corporation, Los Angeles, California

(Received November 20, 1945)

A method applicable to determining droplet sizes in the interior of an already existing atmosphere of fog or mist is developed and described. The method makes use of the fact that if droplets strike a suitably coated surface without wetting the surface, a track of the contact area is formed. A criterion is derived and experimentally evaluated for relating the droplet diameter and the track diameter. An apparatus for utilizing this method is briefly described.

IF a liquid droplet strikes a plane surface without wetting the surface, the droplet will undergo a deformation and momentarily, at least, will have a finite area of contact with the surface. If the surface had been previously coated with, for example, a thin layer of soot, the area of contact of the droplet would be indicated by a disturbance in the soot film which would be clearly visible under a microscope. The carbon particles of the soot have the advantage of being very small and are not easily wetted.

In order to utilize the above, it is necessary to have the relationship between the size of the soot disturbance and the size of the droplet producing the disturbance. As the basis of an analysis for determining this relationship the assumption is made that the only significant forces brought into play during the impact are the inertia forces and the forces due to the surface tension of the droplet. Thus, by this assumption, it is implied that such forces as those due to viscosity, gravity, and compressi-

bility are of negligible importance. Therefore, since

$$F_I \text{ varies as } \rho d^2 v^2,$$

and

$$F_s \text{ varies as } \sigma d,$$

it follows that

$$F_I/F_s \text{ varies as } \rho d v^2 / \sigma.$$

Hence, if only the above assumed forces are important in producing deformation of the droplet during its impact with the soot-coated surface, then the last expression, which is known as Weber's number, is a criterion of dynamical similarity for the phenomenon. Consequently, all dimensionless ratios of quantities related to the phenomenon are some function of the above Weber's number. Thus, the ratio of the track diameter to the droplet diameter can be indicated mathematically by

$$\frac{d'}{d} = \varphi(\rho d v^2 / \sigma).$$

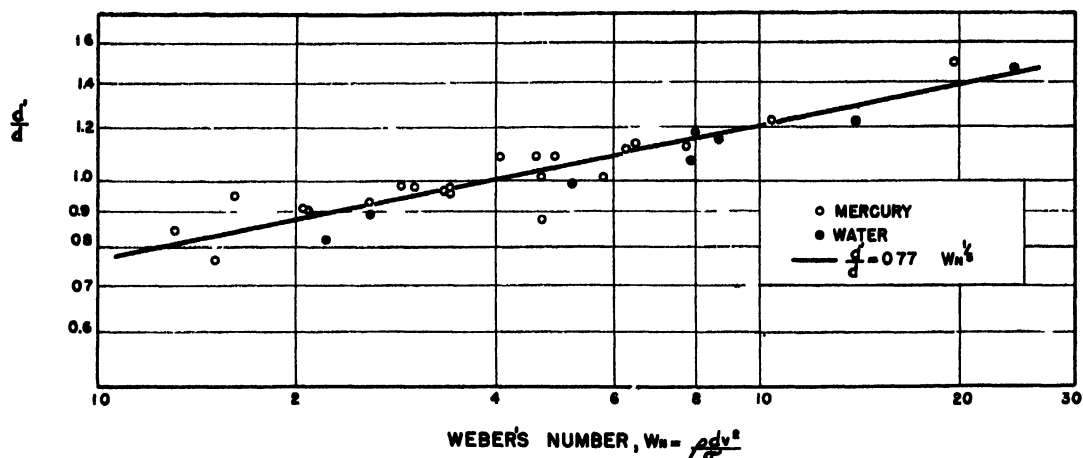


FIG. 1. Experimental relationship between Weber's number and the ratio of track diameter to droplet diameter.

This functional relationship of the diameter ratio and Weber's number was determined experimentally by allowing droplets of mercury and water of known diameters to fall from various measured heights onto a soot-coated glass surface and observing the tracks under a microscope. The velocity at impact was taken as the free fall velocity. The error in the value of Weber's number due to this choice, which neglected air resistance, did not exceed 3 percent in any test and the data are plotted without this correction.

The diameter of the mercury droplets varied from 280 to 800 microns, and were measured at rest under a microscope. For these diameters the shape of the mercury droplets, while at rest on a flat glass surface, differed imperceptibly from spheres and thus this apparent diameter was taken as the diameter under free fall. The measured mercury droplets were swept off the glass surface and fell a known distance to the soot-coated plate below.

The diameter of the water droplets was computed from the weight of a known number of drops formed by a very fine tipped eye-dropper. Water droplets of about 1500 microns in diameter were obtained. In the weighing of a small number (usually ten) of these droplets it was necessary to make a slight correction for evaporation loss. The height of fall of water droplets was taken as the distance of the droplet center above the soot-coated glass just prior to leaving the eye-dropper. This fall distance was varied between one-half and two inches, and was also the range used with the mercury droplets.

The following graph of the ratio of the track diameter to the droplet diameter *versus* Weber's number was plotted from data which were obtained in the above manner (Fig. 1). All experimental points that were determined are included on the plot. Although reasonable experimental care was taken, no claim to less than ± 2 percent error in the plotted results is justified. Never-

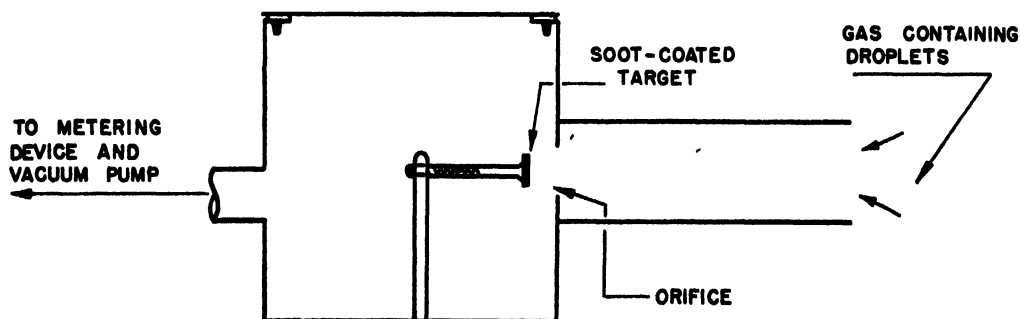


FIG. 2. Schematic sketch of sampling device for obtaining droplet tracks.

theless, it appears that the plotted points define a relationship between the diameter ratio and Weber's number that can be expressed by

$$\frac{d'}{d} = 0.77 \left(\frac{\rho d v^2}{\sigma} \right)^{1/6}.$$

It is apparent from the graph that if Weber's number is about four, the track diameter is equal to the droplet diameter. This would exist, for example, if a water droplet, 5 microns in diameter, struck the coated plate with a velocity of approximately 25 feet per second.

APPLICATION

To apply the above it is more convenient to have the droplet diameter expressed in terms of a Weber's number based on the track diameter. Thus, from the above equation, it can be shown that

$$d = 1.25 \frac{d'}{(\rho d' v^2 / \sigma)^{1/6}}.$$

It is interesting to note that owing to the small fractional exponent in the last equation, the error in the value of the droplet diameter will not exceed one-sixth the error in the measured value of Weber's number.

If the liquid forming the droplets is a pure compound, its density and surface tension will probably be known or can be determined without undue difficulty. If, however, the droplets are not derived from a pure compound, the surface tension and density will depend upon the amount of evaporation that has occurred since their formation. The latter complicates evaluating Weber's number, but in an actual case one is usually able to estimate these variables within reasonable limits and thereby retain fair accuracy for the droplet diameter, since the error in droplet size is much less than the error in Weber's number as was noted above.

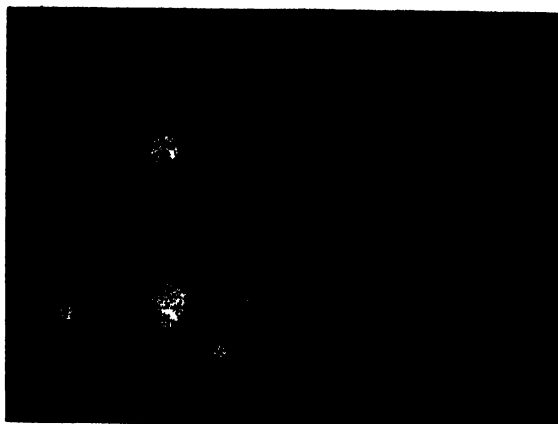


FIG. 3. Photomicrograph (200X) of water droplet tracks on a soot-coated glass target.

The velocity at impact can be determined and controlled by means of a sampling device as indicated in the schematic sketch in Fig. 2. The gas containing the droplets is drawn through a short length of straight pipe and then through an orifice, so that the central portion of the jet impinges against a soot-coated target plate. By metering the gas flow in the suction line and knowing the orifice diameter, the impact velocity can be calculated.

The included photomicrograph, which is an enlargement of 200 diameters, shows water droplet tracks on a portion of a $\frac{1}{4}'' \times \frac{1}{4}''$ soot-coated glass target (Fig. 3). The impingement velocity was approximately 25 feet per second, and the droplet temperature was 60°F.

NOMENCLATURE

- F_I = Inertia force.
- F_σ = Surface tension force.
- ρ = Density of droplet.
- d = Diameter of droplet.
- d' = Track diameter of droplet.
- v = Impingement velocity of droplet.
- σ = Surface tension of droplet.
- W_N = Weber's number.

The Accuracy of Lumping in an Electric Circuit Representing Heat Flow in Cylindrical and Spherical Bodies†

VICTOR PASCHKIS* AND MICHAEL P. HEISLER**

Columbia University, New York, New York

(Received November 29, 1945)

In using lumped resistance-capacitance circuits for studying heat conduction problems the influence of number and size of lumps is important. Several methods of lumping are conceivable for representation of cylindrical or spherical bodies and results of comparative tests show that equal geometrical size of lumps is most accurate. In the various lumps resistance and capacitance have to be in certain definite relationships, which are established in the paper. The influence of number of lumps is also investigated.

FOR the solution of heat flow problems in solids, the electric analogy method has been used increasingly in recent years.¹ The method consists in building a resistance-capacitance circuit representing the bodies subjected to heat flow, impressing on this circuit appropriate boundary conditions, and transforming the measured electrical effects into their thermal counterparts. The application of the electric analogy method is greatly facilitated by the use of a versatile arrangement of resistors, capacitors, switches, and equipment which permits the rapid setting-up of various circuits. One such arrangement is known as the heat and mass flow analyzer.² It consists, briefly, of 15 equal capacitor banks, each bank comprising 6 trays with a total per bank of 41 capacitors of various sizes, between 0.1 and 20 microfarads (total capacitance per bank, 160 microfarads); and of a resistor network comprising a total of 115 sections. Each section has four resistor decades of 10^2 , 10^3 , 10^4 , and 10^5 ohms, respectively. Record-

ing electronic voltmeters allow measurement of 4 voltage-time curves simultaneously. A milli-ampere-minutemeter automatically integrates the current-time curve (equivalent to measuring the total heat flow, not merely the rate of heat flow).

Solutions found on the heat and mass flow analyzer were compared repeatedly with mathematical solutions and direct thermal measurements and found to check closely.³ However, there are inaccuracies inherent in the approximation by lumping. In a previous paper,⁴ the inaccuracy caused by lumping was studied for one particular case: namely, for a circuit representing a slab or large wall of finite thickness but having no end effects, that is, a wall having heat flow in one direction only.

The present paper deals with circuits representing long cylinders and spheres. In such circuits the resistance per unit length increases from the surface towards the center, and the capacitance per unit length decreases.

Because of their importance it is particularly desirable to know how the number of lumps and the way of subdividing the total capacitance and resistance of the circuit influence the results. In a general way the following laws for the radial heat flow in cylinder and sphere can be stated.

Cylinder: The resistance of a cylindrical shell per unit length (in radial direction) is proportional to the logarithm of the ratio of the radii

† This investigation has been made possible through a grant from the Engineering Foundation and has been carried out under the auspices of the A.S.M.E. Committee on Research through the agency of the A.S.M.E. Heat Transfer Division, Subcommittee on Heated and Cooled Enclosures. The authors are indebted to the above organization for its sponsorship and support. Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

* Victor Paschkis is Research Associate in Mechanical Engineering, Columbia University. He is in charge of the technical and scientific work carried out in the Heat and Mass Flow Analyzer Laboratory.

** Michael P. Heisler is Laboratory Assistant in the same laboratory.

¹ V. Paschkis and H. D. Baker, "Method for determining unsteady-state heat transfer by means of an electrical analogy," *Trans. A.S.M.E.* **64**, 105-112 (1942).

² V. Paschkis, "Electrical analogy method for the investigation of transient heat flow problems," *Industrial Heating* **9**, 1162-1170 (1942).

³ M. Avrami and V. Paschkis, "Application of an electrical model to the study of two-dimensional heat flow," *Trans. A.I.C.E.* **38**, 631-652 (June 25, 1942). V. Paschkis, "Establishment of cooling curves of welds by means of electric analogy," *Welding Research Council of The Engineering Foundation*, **8**, 462-483, (October, 1943).

⁴ V. Paschkis and M. P. Heisler, "Accuracy of measurements in lumped R-C cable circuits as used in the study of transient heat flow," *Elec. Eng.* **63**, 165 (1944).

limiting the shell; the capacitance is proportional to the cross-sectional area of the shell.

Sphere: The resistance of a spherical shell is proportional to the difference of the inverse of the inside and outside radii; the capacitance is proportional to the difference of the third powers of the radii limiting the shell.

It must be kept in mind that the capacitance mentioned here is the thermal capacitance, which is proportional to the volume of the body. The electric capacitance is a function of the surface of the body. However, in representing a section of a body, the electric capacitance used to represent its thermal capacitance is proportional to the volume of the lump in the same way as the thermal capacitance.

Obviously the shell at the center has infinite resistance, and part of the work refers to methods of overcoming this difficulty. This characteristic makes it impossible to conceive of a steady-state flow radially through a solid cylinder or sphere, and therefore the influence of capacitor leakage as studied in part of the previous paper⁴ could not be similarly analyzed for the present case.

I. PROBLEMS IN COMMON BETWEEN CYLINDERS AND SPHERES

In working with lumped circuits, a body having evenly distributed electrical or thermal properties—capacitance and resistance—is replaced by a body divided into sections or lumps in which capacitance is concentrated at the center of each lump and resistance along the axis of each lump. This sectionalized body is represented by the electric network of the heat and mass flow analyzer, and determining the influence of the method of lumping on the accuracy of the experimental results was the object of the present investigations.

In slabs there are only two problems to be considered in lumping, the number of lumps and the influence of lumps of unequal thickness. The use of uneven lumps makes it possible to concentrate limited experimental equipment on that part of the circuit which is considered to be the most important (e.g., near the surface, etc.). In each lump, however, the capacitance is always located in the center of the lump. In so arranging the capacitance, the following conditions are fulfilled:

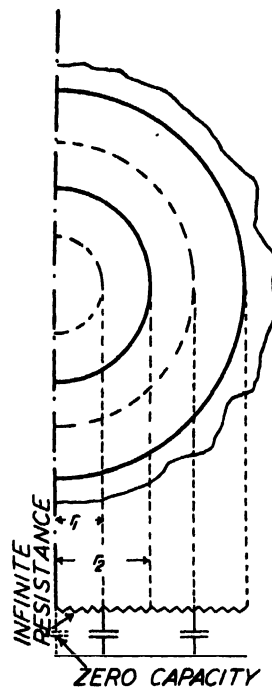


FIG. 1. Explanation of signs for lumping cylinders.

1. Resistance and capacitance both are symmetric in reference to the center.
2. The product of resistance $\times$ capacitance on one side of the center is equal to the product of resistance $\times$ capacitance on the other side.

In cylinders and spheres two additional difficulties arise: First, if the entire radius be divided, either equally or unequally, into a number of lumps, at what points in the lumps should the capacitors be placed? (See Fig. 1.) In Fig. 1, the capacitors are shown in the geometric center of the lump ($r_1 = r_2/2$). However, at least three other positions are defensible: placing the capacitor so that the capacitance on one side of the lump is equal to the capacitance on the other side; placing the capacitor so that the resistances on either side of the capacitor are equal; or putting the capacitor so that the values of the products of resistance $\times$ capacitance are the same for either side of the lump. Second, whatever way of lumping is selected, the first section next to the center offers additional problems. The center of the body has no capacitance, and the resistor connecting the center with the first capacitor is infinitely large, independent of the position of the condenser in the lump. This position, then, determines the resistance of the outside part of

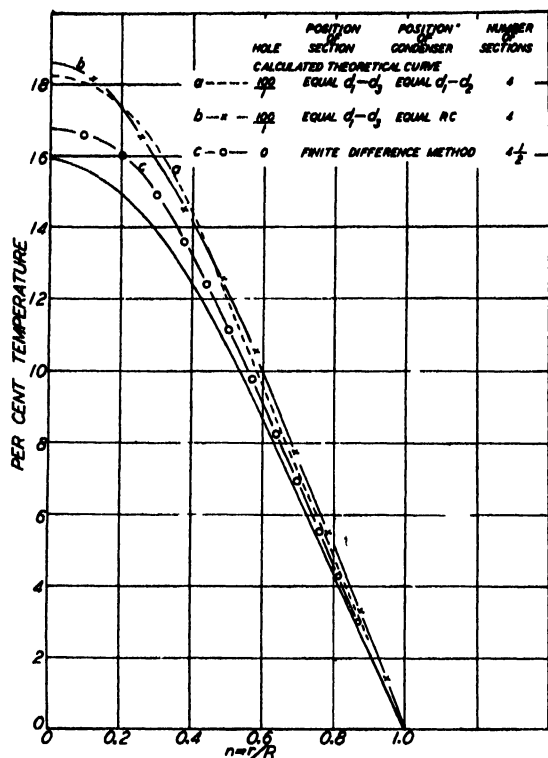


FIG. 2. Cylinder. Comparison of theoretical results (solid line) with experimental results (curves a, b, c), for $at/r^2=0.4$ and four sections. "Equal d_1-d_2 " indicates that the sections have equal thickness. "Equal d_1-d_2 " indicates that the condensers are in the geometric center of each lump. "Equal RC" indicates that the products of resistance $\times$ capacitance on either side of the condenser in each section are equal.

the lump (r_1 to r_3 —Fig. 1) and a minor change in position of capacitor C_1 has a very large influence on this resistance.

One way of overcoming the latter difficulty is to replace the solid cylinder or sphere by a hollow cylinder or sphere having a very small center hole. As soon as a center hole is provided, the resistance from this hole to the innermost condenser becomes finite.

Experiments in which hollow cylinders were substituted for solid ones have revealed various shortcomings in this method. The use of such hypothetical hollow bodies can be avoided, however, by a method of lumping based on finite difference equations. The mathematical basis for this procedure, as related to cylinders, was first published by Jackson and co-workers.⁵ The basis

⁵ R. Jackson, R. J. Sarjant, J. B. Wagstaff, N. R. Eyres, D. R. Hartree, and J. Ingham, "Variable Heat Flow in Steel," paper No. 15 (1944), Iron and Steel Institute, London, England.

for lumping of cylinders by the finite difference method is explained in Appendix I. This method also proved helpful in a similar development for the sphere. The basis for the lumping of spheres is explained in Appendix II. However, the method of lumping and problems of accuracy are numerically different for cylinder and sphere and will therefore be treated separately.

In heating or cooling a body from its surface, the resistance to heat transfer between the surrounding and the surface enters the picture. It has been explained previously⁴ that the severest test for an approximate solution is found in assuming that the surface of the uniformly heated body is subjected to a sudden change of temperature. Therefore, the experiments were carried out for this condition.

All tests have been carried out so that voltage-time curves obtained with differently lumped circuits were compared with theoretically obtained curves.

II. CYLINDER

In working with the cylinder, the feasibility of assuming a hollow cylinder was first analyzed.

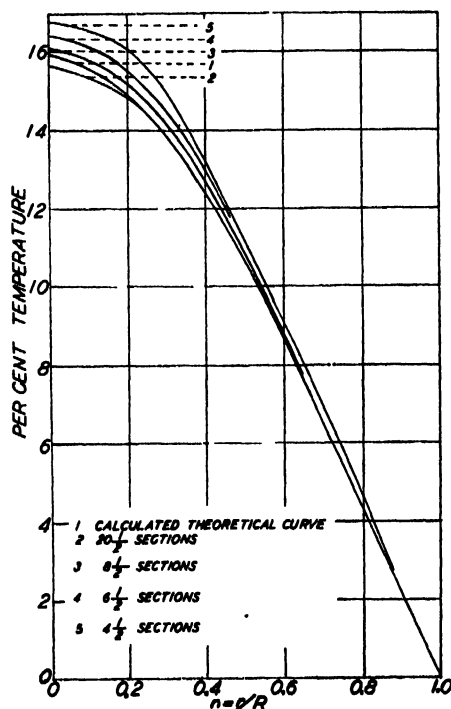


FIG. 3. Cylinder. Comparison of theoretical results (line 1) with experimental results (curves 2-5). Finite difference method; $at/r^2=0.4$; influence of number of sections.

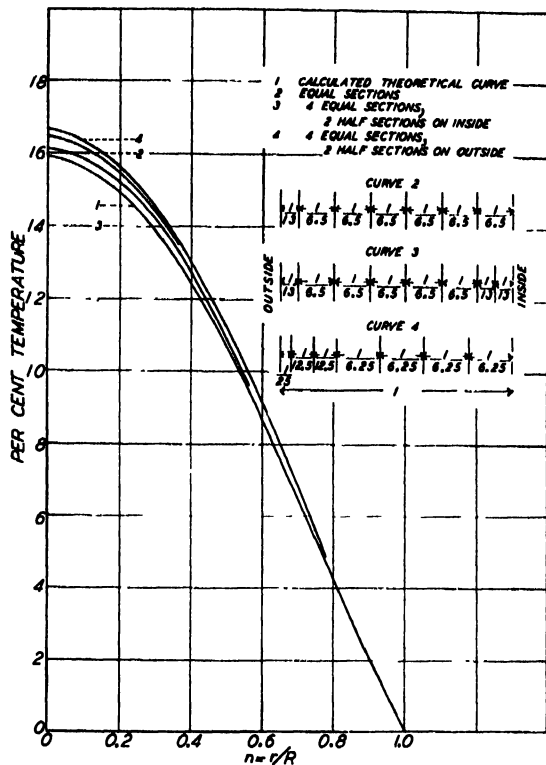


FIG. 4. Comparison of theoretical results (line 1) with experimental results (curves 2-4); $at/r^2=0.4$; finite difference method; $6\frac{1}{2}$ sections; equal and unequal lumping.

In order to verify experimental results it was, of course, necessary to have the theoretical curve. The curves published in McAdams' book on heat transmission (Gurney-Lurie charts) are not very accurate; therefore curves were calculated for $at/r^2=0.2, 0.4$, and 0.6 (a =thermal diffusivity, t =time, r =radial distance). Each of these three curves gives temperatures, expressed in dimensionless units, plotted against relative position " n " in the cylinder. The calculation for the point at the axis of the cylinder was checked against tables published by Schultz and Olsen,⁶ and all points were compared with the values of T. F. Russell.⁷

The greatest number of these investigations was carried out for $at/r^2=0.4$. Figs. 2-4 refer to this dimensionless time. Figure 2 compares different positions of the capacitors for one small assumed hole; the hole was 0.01 times the out-

side diameter. Therefore, the material neglected because of the hole is only 10^{-4} times the entire material of the cylinder. The lumps were selected geometrically, that is, the distance from the outside diameter to the assumed hole diameter was divided into four equal parts. In each part, the capacitor was placed first at the geometric center of the lump and then at a point such that the products of resistance times capacitance on either side of the capacitor were equal. The results are shown by curves a and b . In addition, a third test was run; the lumping in this case was based on finite difference equations; the results are shown by curve c . It is in the nature of this method that there is never a full number of lumps. Therefore, the two curves for the hollow cylinder with four lumps compare with the curve for $4\frac{1}{2}$ lumps for the finite difference method. From Fig. 2 it can be seen that near the center the differences are quite appreciable and reach values more than 15 percent above the calculated values in the most unfavorable case. The two ways of positioning the condenser have no marked influence on the accuracy, but the method of finite differences is considerably better and is accurate to within $5\frac{1}{2}$ percent at the center. Toward the outside all curves more nearly coin-

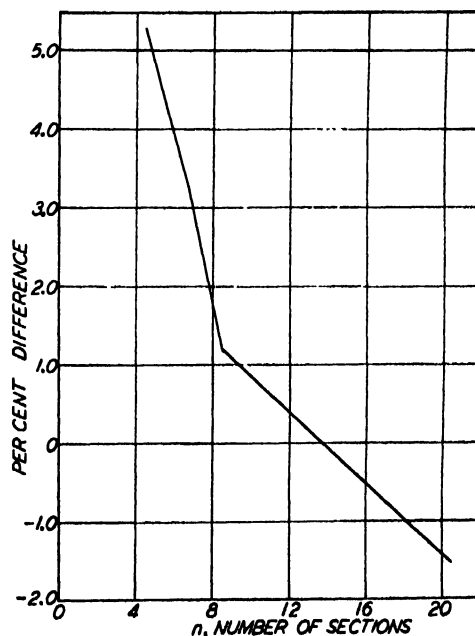


FIG. 5. Cylinder. Percent difference of temperature at center between theoretical curve and experimental curves; $at/r^2=0.4$; finite difference method.

⁶ F. C. W. Olsen and O. T. Schultz, Ind. Eng. Chem. 34, 874 (1942).

⁷ T. F. Russell, "Some Mathematical Considerations on the Heating and Cooling of Steel," special paper No. 14 (1936), Iron and Steel Institute, London, England.

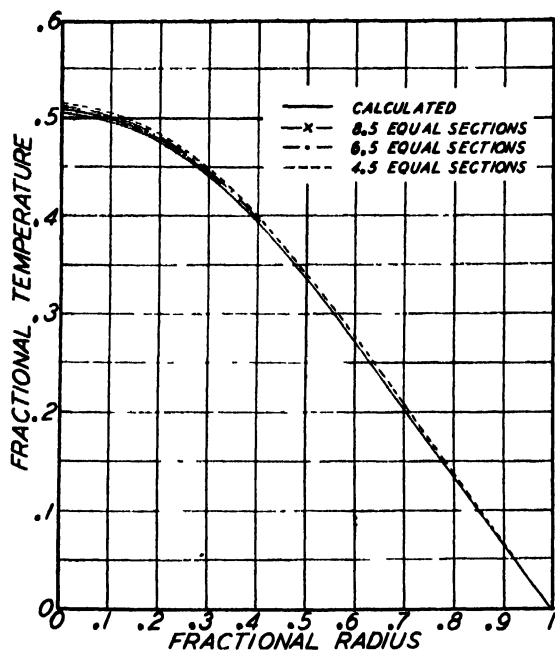


FIG. 6. Cylinder. Finite difference method. Temperature distribution curves with various numbers of sections compared with theoretical curve; $at/r^2 = 0.2$.

cide. However, on a percentage basis the error seems to be almost constant and independent of the distance from the center.

Having found the finite difference method to be satisfactory even with relatively few lumps, all further work was concentrated on this method.

Figure 3 shows the influence of the number of lumps. It is apparent that with $8\frac{1}{2}$ lumps the results are extremely good and will be in error in the order of magnitude of only 1 or $1\frac{1}{2}$ percent. An increase of number of lumps beyond $8\frac{1}{2}$ does not seem to have much influence very close to the center; with $20\frac{1}{2}$ lumps the error becomes negative but of the same order of magnitude as with $8\frac{1}{2}$ lumps. However, at a radius of approximately two-tenths of the outside radius the error with $20\frac{1}{2}$ lumps becomes zero. The influence of the number of lumps on the error is also shown in Fig. 5, where the difference in dimensionless temperature units, expressed as a percentage of the theoretical value, is plotted versus the number of sections.

In many instances it is desirable to have uneven sizes of lumps in order to concentrate available equipment at important points. Figure 4

refers to such uneven lumping conditions. The actual lumping is shown in the sketch on the side. The total number of sections was always $6\frac{1}{2}$. In one case the sections were of equal size (curve 2); in the other instances four larger sections were used together with two sections each one-half of the size of the large sections. In the tests of curve 3 the small sections were on the inside, and in the test of curve 4 they were on the outside. The differences between the three curves are so small as to be considered almost within measuring accuracy. It appears, though, that the small sections on the inside improve the accuracy somewhat; this is not surprising, because it is there that the resistance ratios differ most from the logarithmic theoretical values. In Figs. 6 and 7, the influence of lumping for dimensionless times other than 0.4 were investigated.

Figure 6 shows the curves for $at/r^2 = 0.2$, and it can be seen that the discrepancies between any of the three curves and the theoretical curve are very small indeed. All curves refer to the finite difference method and are distinguished by the number of sections used ($4\frac{1}{2}$, $6\frac{1}{2}$, and $8\frac{1}{2}$ sections). Figure 7 refers to $at/r^2 = 0.6$ and compares the theoretical curve with the curve for finite differences, using $4\frac{1}{2}$ sections. The error is again in the order of magnitude of 5 percent at the center, the same as had been found for $at/r^2 = 0.4$.

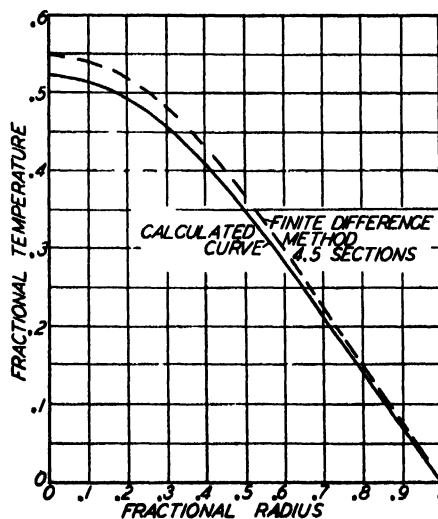


FIG. 7. Cylinder. Finite difference method. Temperature distribution curve for 4.5 sections compared with theoretical curve; $at/r^2 = 0.6$.

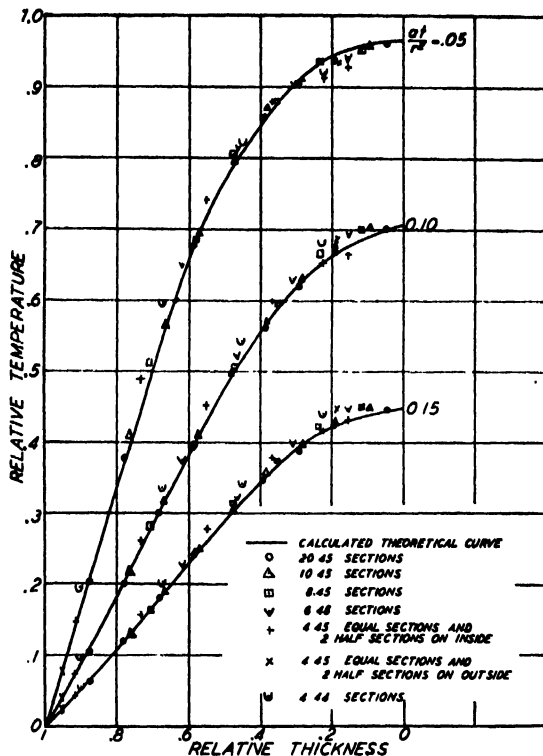


FIG. 8. Spheres. Temperature distribution curves. Relative thickness zero at center; relative thickness 1 at surface. Theoretical curves (full lines) are compared with experimental results by finite difference methods and various numbers of sections as explained in legend.

In conclusion it can be said that for cylinders the finite difference method yields satisfactory results, if 8.5 lumps and more are used.

III. SPHERES

An investigation (Appendix II) shows that for spheres the finite difference method, based on equally sized lumps, still leads to the concept of a hollow sphere—a hole in the center. The capacity in the center turns out to be zero, and the resistance from the center to the next capacitance becomes infinite. If, therefore, the temperatures near the center are of interest, it is necessary to use many lumps. With 10.45 lumps (the fraction 45/100 is explained in the appendix), the point at $0.45/10.45 = 0.041 \times \text{diameter}$ is the one closest to the center. Temperatures closer to the center may be found by extrapolating from the point $0.041 \times \text{diameter}$ and the next point farther out, that is, point $0.041 + 1/10.45 = 0.137 \times \text{diameter}$. If such extrapola-

tion is not sufficient, more lumps may be necessary. For example, with 20.45 lumps, temperatures at $0.45/20.45 = 0.021 \times \text{diameter}$ can be measured directly.

In Fig. 8, results obtained with various numbers of lumps are compared with the theoretical curves; relative thickness is plotted as abscissa and relative temperatures as ordinate. Three theoretical curves are shown, one each for $t/RC (= at/r^2)$ equal to 0.05, 0.10, and 0.15. The results obtained with various methods of lumping are shown grouped about each curve.

With 8.45 lumps, the greatest observed discrepancy is 0.012 of the total ordinate scale of one. In cooling ($t/RC = 0.15$) this causes an error of $0.012/0.43 = 2.8$ percent near the center. Other times and other positions are better. Close to the center there is an error even with 10.45 sections; but with 20.45 sections no dis-

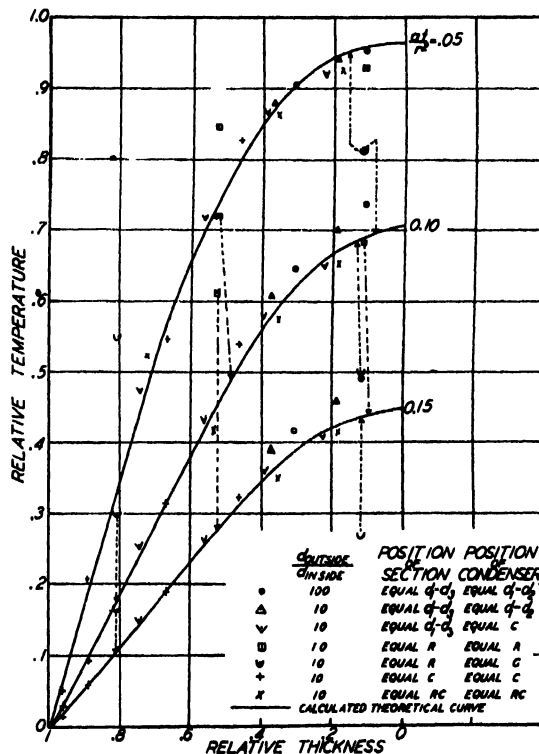


FIG. 9. Spheres. Temperature distribution curves. Relative thickness 1 at surface. Theoretical curves (full lines) are compared with experimental results following different methods as explained by legend. d_{outside} is the outside diameter of the sphere; d_{inside} the diameter of the hole in the center. "Equal $d_1 - d_2$ " indicates that the sections have equal thickness. "Equal C" indicates that the capacitances on either side of the condenser are equal. "Equal R" indicates that the resistances on either side of the condenser are equal.

crepancy, even close to the center, can be detected.

The influence of non-uniform lumping was next examined. If a limited number of lumps is to be used and results are desired particularly in one region, then it may appear useful to make the lumps unequal in size. Thus, two cases were investigated, and in each a total of 7 condensers (6.45 sections) was used. Instead of making each section 1/6.45 of the diameter, the size of each large section was made 1/5.45 of the diameter. One of the large sections was then divided into two smaller sections. In one experiment these smaller sections were located near the center, and in another experiment, near the surface. At early times ($t/RC=0.05$) the splitting of the section near the center does not seem to give very good results, but with increasing time conditions become better. Subdividing on the outside (near the surface) gives excellent results there, but the results near the center are not much better than those obtained with the 4.45 sections.

In Fig. 9, results for lumping based on equal capacitance, equal resistance, etc., are compared with theoretical curves. In no case are results as close as those obtained by the finite difference method.

In conclusion it can be said that also for spheres the finite difference method yields satisfactory results; at least 6.45, preferably 8.45 sections are recommended.

APPENDIX I. FINITE DIFFERENCE METHOD FOR DETERMINING RESISTANCES AND CAPACITANCES IN EVENLY LUMPED CYLINDERS

Figure 10 is a vertical section through a cylinder or sphere and shows (in the lower part of the figure) the wiring diagram of the equivalent circuit. In this case 6 condensers are shown, which would make $m+1=5$. In the figure R stands for resistance, C for capacitance, and V for the potential at any one instant.

The conductance $S=1/R$, time is denoted by t . By applying Kirchhoff's law to point m for $m \neq 0$, it follows that

$$(V_{m+1} - V_m) \cdot S_{m+1} = (V_m - V_{m-1}) \cdot S_m + C_m \cdot dV_m/dt. \quad (1)$$

For the center the equation becomes

$$S_1(V_1 - V_0) = C_0 dV_0/dt. \quad (2)$$

Now, by way of approximation (using a concept originally introduced by Binder and E. Schmidt and applied to cylinders by R. Jackson and co-authors,⁵ and apparently independently by Perry and Berggren⁶), the following statements can be made, referring to a cylindrical shell other than at the center. (See Fig. 11.)

The temperature increase ($Q-S$) at radius $m\Delta r$ can be calculated from:

$$\frac{\partial V_m}{\partial t} = \frac{1}{\rho c 2m(\Delta r)^2} [(2m+1)V_{m+1} - 4mV_m + (2m-1)V_{m-1}] = \frac{\Delta V_m}{\Delta t}, \quad (3)$$

where ρc is the volumetric specific heat.

Now, for the center of the cylinder slightly other conditions prevail, because for $m=0$ the line $P-Q-U$ (Fig. 11) changes direction at Q .

For the center (3) becomes:

$$\frac{\Delta V_0}{\Delta t} = \frac{4}{\rho c (\Delta r)^2} (V_1 - V_0). \quad (4)$$

Equations (1) and (3) are equivalent, if

$$\frac{S_{m+1}}{2m+1} = \frac{S_m + S_{m+1}}{4m} = \frac{S_m}{2m-1} = \frac{C_m}{2m(\Delta r)^2 \rho c}. \quad (5)$$

From (5) it follows that the resistors in the various sections must be in the ratio

$$R_m/R_{m+1} = (2m+1)/(2m-1). \quad (6)$$

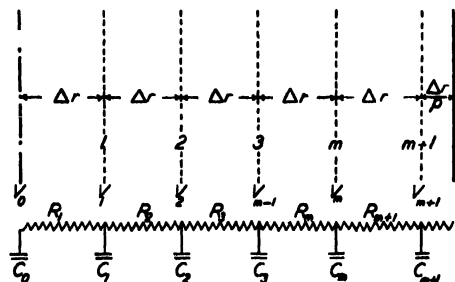


FIG. 10. Explanation of symbols. For finite difference method; cylinders and spheres.

⁵ R. L. Perry, and W. P. Berggren, "Transient heat conduction in hollow cylinders after sudden change of inner-surface temperature," *University of California Publications in Engineering* (University of California Press, Berkeley, California, 1944), Vol. 5, No. 3, pp. 59-88.

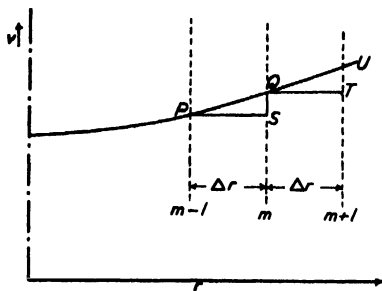


FIG. 11. Method of averaging temperatures in finite difference method.

In the thermal system the diffusivity

$$a \left(\frac{\text{thermal conductivity}}{\text{specific heat} \times \text{density}} \right)$$

takes the place of $1/\rho c$.

If now the thermal times are indicated by T , and the electric times by t , then the time ratio is given by

$$t/T = C_1 R_1 a / 2(\Delta r)^2. \quad (7)$$

APPENDIX II. FINITE DIFFERENCE METHOD FOR DETERMINING RESISTANCES AND CAPACITANCES IN EVENLY LUMPED SPHERES

The procedure for spheres follows closely the one used for cylinders. The approximations expressed by Eqs. (1) and (2) hold for spheres as well as cylinders. For spheres Eq. (3) should be replaced by Eq. (8)

$$\frac{\Delta V_m}{\Delta t} = \frac{1}{\rho c m (\Delta r)^2} \times [(m+1)V_{m+1} - 2mV_m + (m-1)V_{m-1}]. \quad (8)$$

This equation is found by transforming the basic equation

$$\frac{\partial V}{\partial t} = \frac{1}{\rho c} \left(\frac{\partial^2 V}{\partial r^2} + \frac{2}{r} \frac{\partial V}{\partial r} \right). \quad (9)$$

The rate of change is approximately

$$\begin{aligned} \frac{\Delta^2 V}{\Delta r^2} &= \left(\frac{UT}{\Delta r} - \frac{QS}{\Delta r} \right) \frac{1}{\Delta r} \\ &= \frac{(V_{m+1} - V_m) - (V_m - V_{m-1})}{(\Delta r)^2} \\ &= \frac{V_{m+1} - 2V_m + V_{m-1}}{(\Delta r)^2}. \quad (10) \end{aligned}$$

The approximate value of $\partial V_m / \partial r$ can be found by replacing the curve PU by a straight line; then:

$$\Delta V_m = (V_{m+1} - V_{m-1}) / 2. \quad (11)$$

Finally $r = m\Delta r$. By combining these equations with Eq. (9), Eq. (8) may be obtained. Equations (1) and (8) are equivalent, if

$$\frac{S_{m+1}}{m+1} = \frac{S_m + S_{m+1}}{2m} = \frac{S_m}{m-1} = \frac{C_m}{\rho c m (\Delta r)^2}. \quad (13)$$

From (13) it follows that the resistors in the various sections must be in the ratio

$$R_m / R_{m-1} = (m+1) / (m-1). \quad (13a)$$

It can be readily seen that, for $m=1$,

$$R_m / R_{m+1} = \infty.$$

Hence the inside resistance becomes infinite, and as will be seen later, C_0 becomes 0. Therefore it is convenient to call $R_2 = 1$. Then

$$R_2 : R_3 : R_4 : R_5 : \dots = 1 : \frac{1}{3} : \frac{1}{6} : \frac{1}{10} \dots$$

It can be readily seen that the series is formed by fractions, the numerators of which always equal one. The denominators increase by adding 2, 3, 4, 5, etc.; the first denominator is 1, the next 3=1+2, then 6=3+3, then 10=6+4, then 15=10+5, etc. The resistance of a spherical shell is proportional to $(1/d_i) - (1/d_o)$, where d_i and d_o are the inside and outside diameters, respectively. The ratio of resistance of the second to that of the third shell is

$$\frac{1/\Delta r - 1/2\Delta r}{1/2\Delta r - 1/3\Delta r} = \frac{\frac{1}{2} - \frac{1}{3}}{\frac{1}{3} - \frac{1}{4}} = \frac{3}{1},$$

which is exactly the ratio found from Eq. (13a).

From Eq. (13), it follows that the capacitors other than the one in the center must be in the ratio:

$$C_1 : C_2 : C_3 : C_4 : \dots = 1 : 4 : 9 : 16 : \dots$$

Actually the capacitors should be proportional to the volumes of the shells, and should be located at the center of the shell. The question arises where the end of each section is. If the capacitor is to be at the mean radius, then the sectioning would be at $0.5r$; $1.5r$; $2.5r$; etc.

The volume of each shell is proportional to the

TABLE I.

1	2	3	4	5	6	7
Section No.	Radius $m\Delta r$	$(m\Delta r)^3$	Difference $(m\Delta r)^3$	Volume ratio	Volume ratio from Eq. (14)	Comparison of volume ratios Col. 5/Col. 6
1	.5	0.125				
2	1.5	3.375	3.25	1	1	1
3	2.5	15.625	12.25	3.77	4	0.942
4	3.5	42.875	27.25	8.39	9	0.933
:	:	:	:	:	:	:
:	:	:	:	:	:	:
11	10.5	1157.625				
12	11.5	1520.875	363.25	111.8	121	0.923

difference of the third powers of the outside and inside radius. Table I shows for some sections the radius, the third power of the radius, the difference of the third powers, the ratio based on these differences, the ratio as found from Eq. (13) and a comparison between the two ratios.

Because the comparison of the volume ratios (Col. 6) does not appear to be very good, the capacitor was put on a radius other than the mean value. The results with the radii at 0.45, 1.45, 2.45, etc., appear to be better, as may be seen from Table II. In order to determine the value of the center capacitor, C_0 , equations

similar to (4) must be found. By writing Eqs. (10)–(12) for the center the expression (8a) replacing Eq. (4) is found.

$$\frac{\Delta V_0}{\Delta t} = \frac{6}{\rho c(\Delta r)^2} (V_1 - V_0). \quad (8a)$$

TABLE II.

1	2	3	4	5	6	7
Section No.	Radius $m\Delta r$	$(m\Delta r)^3$	Difference $(m\Delta r)^3$	Volume ratio	Volume ratio from Eq. (14)	Comparison of volume ratios Col. 5/Col. 6
1	.45	.091				
2	1.45	3.049	2.958	1	1	
3	2.45	14.706	11.657	3.94	4	0.985
4	3.45	41.063	26.357	8.91	9	0.992
.	.	.				
.	.	.				
11	10.45	1141.17				
12	11.45	1501.12	359.95	121.6	121	1.003

In order to determine the value of the center capacitor combine Eq. (2) with (8a):

$$C_0 = \rho c(\Delta r)^2 / 6R_1. \quad (14)$$

As stated above the resistance $R_1 = \infty$; hence $C_0 = 0$.

Determination of the Temperature Rise and the Maximum Safe Current through Multiconductor Electric Cables*

HAIG P. ISKENDERIAN

Federal Telephone and Radio Corporation, New York, New York

AND

WILLIAM J. HORVATH

Navy Department, Washington, D. C.

(Received November 29, 1945)

A general method of calculating the temperature rise in multiconductor rubber insulated electric cables based upon the thermal constants of the cable has been derived. The predictions of the theory have been verified experimentally for a number of cables and values of the thermal resistivity constants have been obtained as a function of size and composition of the cable. The results are used to predict the maximum safe current for different size cables to limit the temperature rise at the center to 25°C above an ambient temperature of 50°C.

INTRODUCTION

IN the design of cable installations it is necessary to know the maximum current which can be sent through a given cable without raising its temperature to a value which would injure the insulation. In general the maximum temperature which the insulation will endure under continuous operation is known, and the problem becomes one of calculating the temperature rise in the cable as a function of the current running through it. Previous work on this subject¹ has been confined to cables containing one, two, or three conductors. The present investigation develops the complete theory for a cable consisting of any number of rubber covered electric wires, all carrying the same current.

Application of the theory depends on accurate knowledge of the radial and surface thermal resistivities. These have been obtained with considerable precision by measuring the temperature drops in experimental cables. The rubber insulation used in these cables is known to withstand continuous operation at 75°C. The maximum safe currents have therefore been calculated for a temperature rise at the center of the cable of 25°C over a 50°C ambient.

I. THEORY

The theoretical expressions are derived first for a cable made up of an infinite number of conductors distributed uniformly in a cylindrical insulating medium. In order to apply these expressions to actual cables it is necessary to take into account the variation of the thermal constants of the cable with the size and number of wires. Empirical relations for these constants have been established experimentally. The excellent agreement between theory and experiment is adequate justification for this method of procedure.

The temperature as a function of time may be obtained by writing the fundamental heat relationship for an element of cylindrical shell of unit length and radii x and $x + \Delta x$:

Heat generated = Heat dissipated + Heat stored.

$$I^2 R_{cu} N_{x, x+\Delta x} dt = - \left\{ \frac{2\pi x}{\rho} \left(\frac{dT}{dx} \right)_x - \frac{2\pi(x+\Delta x)}{\rho} \left(\frac{dT}{dx} \right)_{x+\Delta x} \right\} dt + 2\pi x \Delta x S_0 d_0 dT, \quad (1)$$

where

I = current in amperes through each wire,

R_{cu} = resistance in ohms per cm of each conductor,

$N_{x, x+\Delta x}$ = number of conductors in a shell of radii x and $x + \Delta x$,

$$= \frac{2\pi x \Delta x}{\pi r^2} n$$

(r is the radius of the cable in cm; n is the number of conductors in the cable),

* This work was done in the Spring of 1941 while the authors were employed by the Bureau of Ordnance, Navy Department.

¹ Report by S. J. Rosch, published by the N.E.M.A. (January, 1938).

S_0 = average specific heat of cable in watt sec./°C,
 d_0 = average density of cable,
 T = temperature rise in °C, and
 ρ = radial thermal resistivity of the cable in °C/watt/cm

Substituting for $N_{x, x+\Delta x}$ and noting that the second term in (1) is the differential of $(2\pi x/\rho) \cdot (dT/dx)$, we obtain:

$$\frac{dT}{dt} = a + \frac{b}{x} \frac{d}{dx} \left(x \frac{dT}{dx} \right), \quad (2)$$

where,

$$a = \frac{I^2 R_{cu} n}{\pi r^2 S_0 d_0} \quad \text{and} \quad b = \frac{1}{\rho S_0 d_0}. \quad (3)$$

For the steady state $dT/dt = 0$ and (2) becomes by successive integrations:

$$T_x = T_c - \frac{ax^2}{4b}, \quad (4)$$

where T_x = the temperature rise at equilibrium, at a distance x from the center of the cable, and T_c = the temperature rise at equilibrium at the center of the cable.

The general solution of (2) must satisfy the following initial and boundary conditions:

$$\left. \begin{array}{l} t=0 \\ T=0 \end{array} \right\} \quad (I) \quad \text{and} \quad T_s = -\frac{B}{\rho} \frac{dT}{dx} \Big|_{x=r}, \quad (11)$$

where B is the thermal surface resistivity in °C/watt/cm².

The solution may be expressed in the following form:

$$T(x, t) = f(t)u(x) + T_c - \frac{ax^2}{4b}, \quad (5)$$

where $f(t)$ is a function of time only and $u(x)$ is a function of radial distance only.

By substituting (5) in (2) and by separating variables we obtain:

$$\text{and } \left. \begin{array}{l} f(t) = C_0 \exp(-\alpha^2 bt) \\ \frac{d^2 u}{dx^2} + \frac{1}{x} \frac{du}{dx} + \alpha^2 u = 0 \end{array} \right\}, \quad (6)$$

where C_0 and α are constants.

The second equation in (6) is Bessel's equation.

Thus our solution (5) becomes:

$$T(x, t) = \sum_i A_i J_0(\alpha_i x) \exp(-\alpha_i^2 bt) + T_c - \frac{ax^2}{4b}. \quad (7)$$

The constants are obtained by making use of the boundary condition (II) and Eq. (7):

$$\begin{aligned} \sum_i A_i J_0(\alpha_i r) \exp(-\alpha_i^2 bt) + T_c - \frac{ar^2}{4b} \\ = -\frac{B}{\rho} \left\{ \sum_i A_i \alpha_i J_1(\alpha_i r) \exp(-\alpha_i^2 bt) + \frac{ar}{2b} \right\}. \end{aligned}$$

Separating the variables this expression reduces to

$$J_0(\alpha_i r) - \frac{B}{\rho} \alpha_i J_1(\alpha_i r) = 0 \quad (8)$$

and

$$T_c = \frac{ar^2}{4b} + \frac{B}{\rho} \frac{ar}{2b} = I^2 R_{cu} n \left(\frac{\rho}{4\pi} + \frac{B}{\pi D} \right), \quad (9)$$

where α_i are the roots of Eq. (8).

The relation (9) gives the maximum temperature rise at the center of the cable. It may be written in the form of two equations:

$$T_r = I^2 R_{cu} n \frac{\rho}{4\pi}, \quad (10)$$

$$T_s = I^2 R_{cu} n \frac{B}{\pi D}, \quad (11)$$

where T_r = temperature rise in the cable from $x=r$ to $x=0$ and T_s = temperature rise at the surface of the cable.

From (10) to (11) the constants ρ and B may be found in terms of the radial and surface temperature rises, the power loss, and the diameter of the cable. Or, conversely, knowing ρ , B , D , and $I^2 R_{cu} n$ of a cable, the maximum temperature rise, T_c , may be computed.

To evaluate the coefficients, A_i , use is made of the initial condition (I) and Eq. (7):

$$A_1 J_0(\alpha_1 x) + A_2 J_0(\alpha_2 x) + \dots, \text{ etc.} = -T_c + \frac{ax^2}{4b},$$

which yields by virtue of the orthogonality

properties of the Bessel functions:

$$A_1 = \frac{2\alpha_1^2}{\left[\left(\frac{\rho}{B}\right)^2 + \alpha_1^2 r^2\right] [J_0(\alpha_1 r)]^2} \times \int_0^r x \left(-T_c + \frac{ax^2}{4b}\right) J_0(\alpha_1 x) dx; \quad (12)$$

carrying out the integration and substituting for a and b , we find for A_1 :

$$A_1 = \frac{2\alpha_1^2}{r^2 \left[\alpha^2 + \left(\frac{\rho}{B}\right)^2\right] [J_0(\alpha_1 r)]^2} \left\{ \frac{I^2 R_{cu} n \rho}{2\pi \alpha_1^2} J_0(\alpha_1 r) + \left[\frac{I^2 R_{cu} n \rho}{4\pi \alpha_1^2} \left(\alpha_1 r - \frac{4}{\alpha_1 r}\right) - \frac{T_c r}{\alpha_1} \right] J_1(\alpha_1 r) \right\}. \quad (13)$$

Having determined the constants A_1 and α_1 of Eq. (7), we now have the complete solution for the temperature of the cable as a function of time and radial distance.

The time variation of temperature at the center of the cable is of special interest. For this case Eq. (7) reduces to:

$$T(0, t) = A_1 \exp(-\alpha_1^2 b t) + A_2 \exp(-\alpha_2^2 b t) + \dots, \text{etc.} + T_c. \quad (14)$$

It will be shown in Part III that the coefficients, A_i , converge very rapidly and therefore the above expression reduces to the form:

$$T(0, t) \simeq T_c [1 - \exp(-\alpha_1^2 b t)]. \quad (15)$$

A comparison between Eq. (15) and the experimental curves is made in Part III.

II. EXPERIMENTAL PROCEDURE

After considerable experimentation with thermocouple methods, it was found that the measurement of the change of resistance of the copper conductors gave much more reproducible results and consequently all measurements were made by this method.

If the resistance is measured at some known temperature T_0 and again at some other temperature T_1 , then we may use the relationship

$$R_{T_1} = R_{T_0} [1 + \beta(T_1 - T_0)]$$

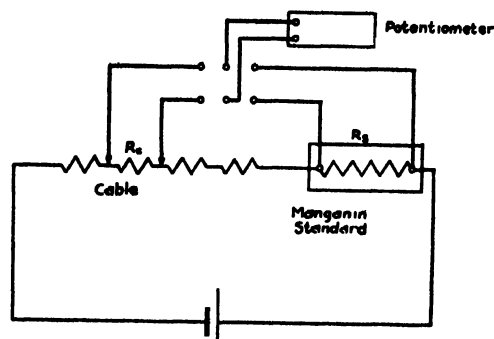


FIG. 1.

to obtain the temperature difference $T_1 - T_0$

$$T_1 - T_0 = \frac{1}{\beta} \left[\frac{R_{T_1}}{R_{T_0}} - 1 \right], \quad (16)$$

where R_{T_0} is the resistance at a known temperature, T_0 , R_{T_1} is the resistance at any other temperature, T_1 , and β is the temperature coefficient of resistivity of copper $= 1/(234.5 + T_0)$.

The circuit used for the measurements is shown schematically in Fig. 1. A standard manganin resistor, R_s , is placed in series with the cable so that all resistance measurements could be referred to this standard. Let R_{c,T_0} be the resistance of one of the wires in the cable at the initial temperature T_0 . The ratio $R_{c,T_1}/R_{c,T_0}$ in Eq. (16) may be obtained by voltage measurements as follows:

From Ohm's law,

$$V_{c,T_0} = I T_0 R_{c,T_0}$$

and

$$V_{s,T_0} = I T_0 R_{s,T_0};$$

hence,

$$R_{c,T_0} = \frac{V_{c,T_0}}{V_{s,T_0}} R_{s,T_0}, \quad (17)$$

where V , I , and R have their usual meanings, the subscript, T_0 , referring to the values of these quantities when the cable is at T_0 degrees centigrade, and the subscripts, c and s , to the cable and manganin resistance, respectively.

Similarly

$$R_{c,T_1} = \frac{V_{c,T_1}}{V_{s,T_1}} R_{s,T_1}. \quad (18)$$

Since the temperature coefficient of manganin is negligibly small,

$$R_{s,T_0} = R_{s,T_1};$$

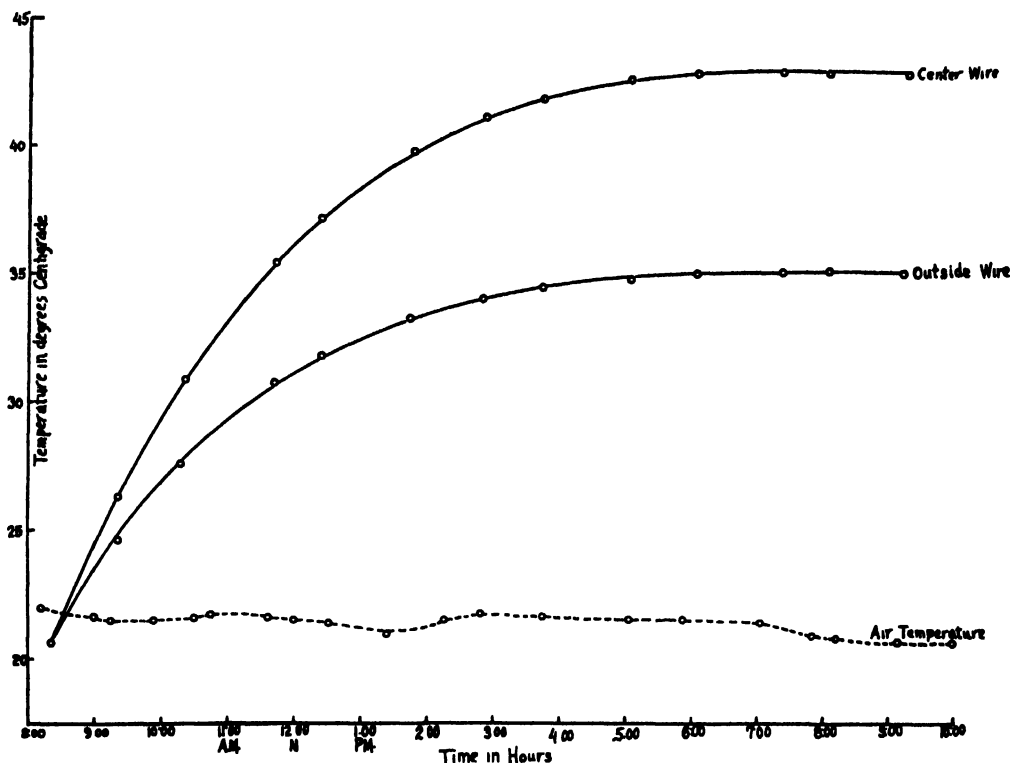


FIG. 2. Experimental variation of cable temperature in a cable having 61 conductors of No. 3 wire when a current of 27 amperes flows continuously through it.

dividing (18) by (17) and substituting the result in (16) yields:

$$T_1 - T_0 = - \frac{1}{\beta} \left[\frac{V_c T_1}{V_s T_1} \frac{V_s T_0}{V_c T_0} - 1 \right]. \quad (19)$$

The expression (19) gives the temperature rise in terms of measurable quantities. The voltage

TABLE I. Experimental data on cables

No. of wires in cables	AWG No.	No. of layers in cables	Dimensions of cables		Current in amp.	Total power in watts	T_s temp. rise at center in °C	Radial temp. rise in °C
			Length (cm)	Diameter (cm)				
19	1	2	375	7.24	57	104	20.5	4.0
37	1	3	416	10.04	45	135	21.0	5.7
61	1	4	415	14.06	37	162	21.8	8.1
19	3	2	412	6.48	42	100.7	19.5	4.3
37	3	3	417	8.80	33	120	20.9	6.1
61	3	4	415	11.20	27	134	21.4	7.7
91	3	5	420	13.12	24	160	23.5	9.5
19	6	2	414	6.06	27	82.0	17.0	4.1
37	6	3	420	8.12	22	107.5	20.0	6.0
61	6	4	416	9.52	18	115.3	20.9	7.3
19	9	2	424	4.44	17	77.0	18.1	4.1
37	9	3	421	5.90	14	85.5	18.7	5.3
61	9	4	420	7.78	12	97.0	19.6	6.5
37	15	3	425	4.14	5.8	58.2	14.9	3.6
61	15	4	422	5.20	5	71.5	16.6	4.8
61	18	4	425	4.80	4	96.8	25.3	6.6

drops across R_c and R_s were measured by means of a Rubicon type "B" potentiometer.

The cables used in the experiment are listed in Table I. All were made up out of standard rubber covered wire which had a safe operating limit of 75°C. The cables were assembled in such a way that a given wire remained in the same layer throughout the length of the cable. A canvas sleeve was sewn over the assembled cable and this was painted with a coating of preservative paint. Subsequent measurements showed that the painting had no measurable effect on the results of the experiments.

All the wires in the cable were put in series so that the current flowed back and forth through the length of the sample. Each sample was approximately 14 feet long, the ends being enclosed in two wooden boxes packed with insulating material. With this arrangement, thermocouples showed no appreciable temperature gradient along the length of the cable. This was essential since the resistance method yields an average value of the temperature in any one layer and not just the temperature at one point.

Since it would not be practical, especially in the larger cables, to measure the voltage drop across every wire, only four wires in each layer were selected for these measurements. The leads from these wires were run to two selector switches, S_1 and S_2 , and thence to two double pole-double throw switches, S_3 and S_4 , leading to the potentiometer. S_3 served as a reversing switch, and S_4 connected the potentiometer either to the manganin resistor or to one of the wires in the cable.

The current was furnished by a d.c. motor generator set with four storage batteries "floating" across the output. Since the fluctuations in voltage across the cable were large (about $\frac{1}{2}$ percent) a switch, S_5 , was installed which disconnected the generator from the cable, leaving only the batteries to furnish the current during measurements. This switch was thrown about 10 minutes before every reading to allow the batteries to come to equilibrium. With this arrangement there was no difficulty in reading the voltage drops to within .01 millivolt.

A complete run on a cable consisted of the following:

1. The voltage drops V_{c,T_0} and V_{s,T_0} were measured with reduced currents to prevent any heating during the measurement. T_0 was measured by means of several Fe-Con. thermocouples imbedded in the cable. The "cold" junction of these couples was kept in air but its temperature could be read to 0.1°C .

2. Readings were taken of V_{c,T_1} and V_{s,T_1} at hourly intervals using only the center wire and one of the outer wires. With the aid of these readings the approach to equilibrium could be studied.

3. When equilibrium had been reached, the voltage drops across the other wires were measured and the temperatures were calculated by means of (19).

The ambient temperature was read at half-hour intervals on several mercury thermometers located 3 or 4 feet from the cable. The air temperature as measured in different parts of the room agreed to within two-tenths of a degree, indicating a relatively draft-free condition.

The over-all accuracy of the method was estimated to be between 0.2° – 0.3°C .

III. RESULTS OF MEASUREMENTS

Measurements were made on the sixteen cables listed in Table I. The currents which were used were chosen rather arbitrarily to give a temperature rise of 20° – 25°C at the center of the cable. Several runs were usually made on each cable and curves were plotted showing the variation of temperature with time as well as the radial temperature distribution at equilibrium. Typical curves are shown in Figs. 2 and 3. From such curves as in Fig. 3 it is possible to obtain the surface temperature rise, T_s , and the radial temperature rise, T_R . These have been listed for the test cables in Table I.

From the measured values of the temperature rise and the power input, the values of the radial thermal resistivity, ρ , and the surface thermal resistivity, B may be calculated according to Eqs. (10) and (11). The resultant values for ρ are plotted in Figs. 4 and 5 which show the variation of ρ with size of wire and number of layers. Figure 6 shows the variation of B with the diameter D of the cable. The general shape

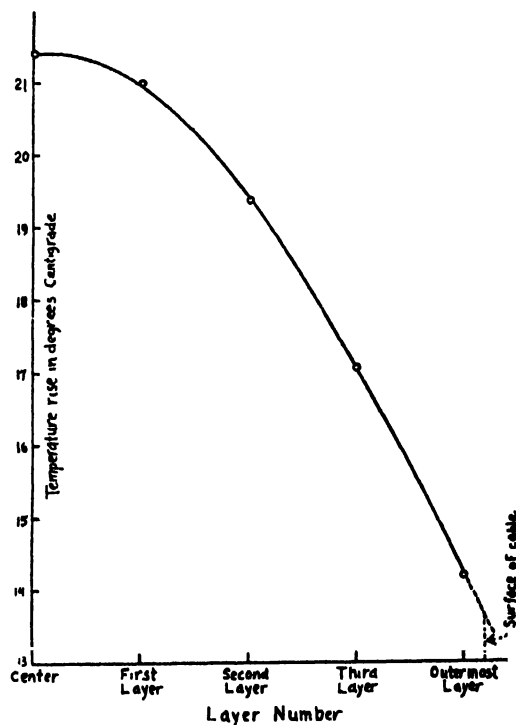


FIG. 3. Temperature rise in each layer of the cable having 61 turns of No. 3 wire (diameter 11.2 cm).

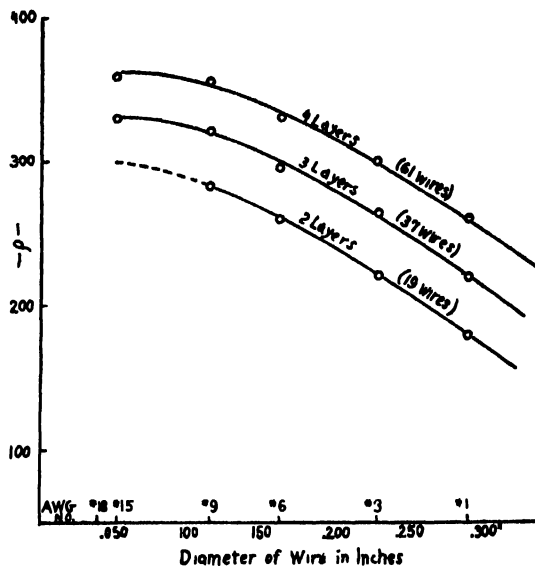


FIG. 4. Value of thermal resistivity, ρ , in degrees centigrade per watt per centimeter, for wire cables.

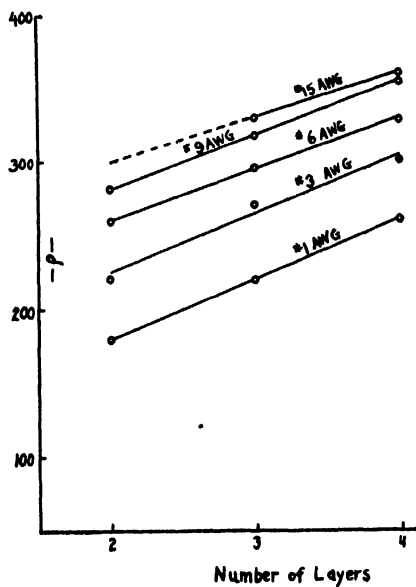


FIG. 5. Values of thermal resistivity, ρ , in degrees centigrade per watt per centimeter, for wire cables.

of this last curve is quite similar to the one obtained by S. J. Rosch.¹

By interpolation of these curves, the constants for all other cables may be obtained and the maximum safe current computed from:

$$I = \left\{ \frac{25}{R_{cu} n [\rho / 4\pi + B / \pi D]} \right\}^{\frac{1}{2}} \quad (20)$$

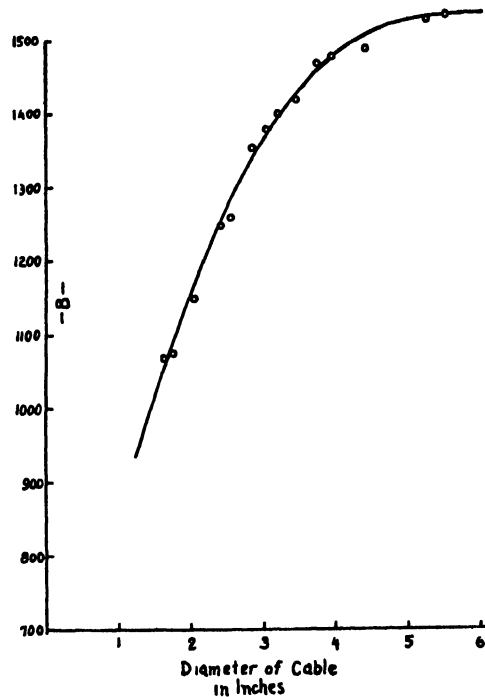


FIG. 6. Values of thermal surface resistivity, B , in degrees centigrade per watt per square centimeter, for wire cables.

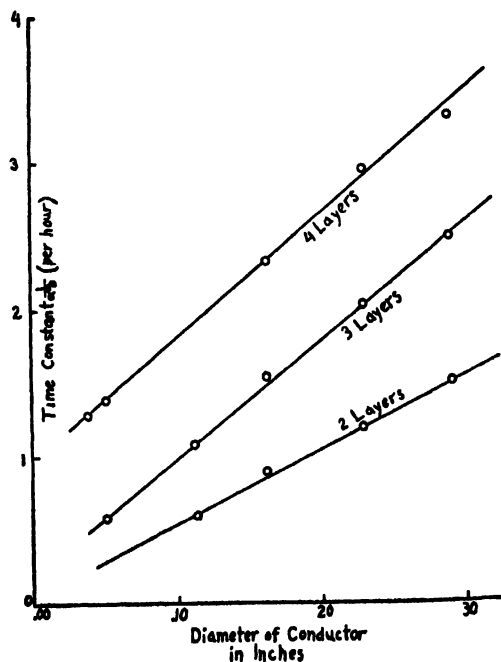


FIG. 7. "Time constant" of cable vs. size of conductor.

TABLE II. Average temperature in cables. For a rise of 25°C in cable and an ambient temperature of 50°C.

Number of layers	2	3	4	5
Average temperature (°C)	71.7	70.7	69.8	68.8

This current will give a 25°C rise in temperature at the center of the cable. Since we wished to allow for an ambient of 50°C, the resistance of the cable per cm R_{cu} had to be calculated for the average temperature of the copper which would exist at that ambient with the maximum current flowing. These average temperatures were found to be independent of the wire size and depended only on the number of layers. The values are listed in Table II. The currents were calculated from Eq. (20) using resistances from standard tables at the average temperature shown in Table II. The results of these calculations are given in Table III for cables with 19, 37, and 61 wires, corresponding to 3, 4, and 5 layers. Intermediate values may be obtained satisfactorily by interpolation.

In order to compare the theoretical expression (14) with the actual time variation of the temperature at the center of the cable, the coefficients A_1 , A_2 , and A_3 were calculated for several cables. They were found to converge very rapidly, and in the worst case (91 conductors of No. 3 wire) A_1 differed from T_0 by only 4 percent. The expression (15) may therefore be used to get the temperature-time relation at the

TABLE III. Maximum safe currents for cables. For a rise of 25°C in cable and an ambient temperature of 50°C.

No. of wires in cables	AWG No.	Amperes	Diameter of cable (inches)	No. of wires in cables	AWG No.	Amperes	Diameter of cable (inches)
19	0	70.8	3.16	19	6	30.9	2.38
37	0	54.4	4.43	37	6	23.4	3.20
61	0	44.6	5.70	61	6	18.7	3.75
19	1	60.0	2.85	19	7	27.0	2.02
37	1	46.7	3.95	37	7	20.4	2.83
61	1	38.0	5.54	61	7	16.6	3.64
19	2	52.9	2.66	19	8	23.0	1.73
37	2	40.2	3.72	37	8	17.5	2.42
61	2	32.5	4.79	61	8	14.2	3.11
19	3	45.0	2.55	19	9	19.1	1.75
37	3	34.2	3.46	37	9	15.3	2.32
61	3	27.8	4.41	61	9	12.9	3.06
91	3	23.7	5.17	19	12	14.1	1.47
19	4	40.2	2.36	37	12	10.6	2.05
37	4	30.4	3.30	61	12	8.6	2.64
61	4	24.8	4.25	37	15	7.1	1.63
19	5	35.2	2.23	61	15	6.4	2.05
37	5	26.6	3.12	61	18	3.8	1.89
61	5	21.6	4.02				

center of the cable:

$$T(0, t) = T_0[1 - \exp(-\alpha_1^2 bt)], \quad (21)$$

where $\alpha_1 = f((B/\rho), r)$ is a root of (8) and the value of b is given by $b = 1/\rho S_0 d_0$. $S_0 d_0$ has been calculated from handbook values for the specific heats and densities of rubber, copper, and air by using the expression

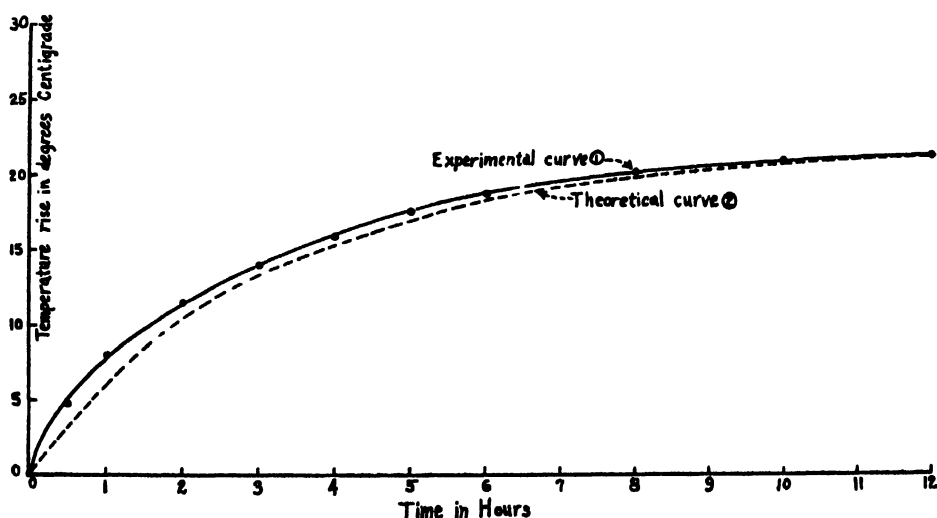


FIG. 8. Variation of temperature with time at the center of a cable having 61 conductors. ①—Experimental curve; ②—Theoretical curve.

$$S_0 d_0 = \frac{1}{A} [(a S d)_{\text{copper}} + (a S d)_{\text{rubber}} + (a S d)_{\text{air in cable}}],$$

A = cross-section area of cable,
 S = specific heat in watt sec./deg. C,
 d = density in g/cm³, and
 a = cross-section area of constituents in cm².

Values of $\alpha_1^2 b$ have been obtained for all the cables tested. The reciprocal of $\alpha_1^2 b$, the time constant, is plotted in Fig. 7 as a function of the diameter of the conductor.

In Fig. 8 the experimental temperature-time

curve for one of the larger cables is compared with the theoretical curve of Eq. (15). The agreement between the two is quite good and indicates that Eq. (15) may be used with considerable confidence to represent the temperature-time relationship at the center of a multiconductor cable.

ACKNOWLEDGMENT

The authors wish to express their thanks to Dr. George I. Welch for many helpful discussions of the problem and to Mr. George Bigelow for assistance with the experimental techniques.

A Computer for Solving Linear Simultaneous Equations

CLIFFORD E. BERRY, DOYLE E. WILCOX, SIBYL M. ROCK, AND H. W. WASHBURN
Consolidated Engineering Corporation, Pasadena, California

(Received November 16, 1945)

The mathematical principles of the classical iterative method of solving linear simultaneous equations are discussed. Basic electrical circuits for setting up an analog of the mathematical relations are given, and a commercial model of a 12-equation computer is briefly described. The results of solving a number of problems on the computer are given to illustrate its accuracy and speed of operation. It is found that solving sets of 12 equations requires only 1/4 to 1/7 the time required by conventional methods.

DURING the war a large number of mass spectrometers and infra-red spectrophotometers were installed in petroleum refineries and research institutions. In the application of each of these instruments to analytical work there is the frequent necessity of solving linear simultaneous algebraic equations, particularly in analysis of complex hydrocarbon mixtures. Often mixtures requiring the solution of eight or ten simultaneous equations for their analysis are encountered, and occasional mixtures may involve as many as twelve equations. As is well known, the labor involved in solving sets of this size is considerable, with the result that the present limitation on speed of analysis is often not the obtaining of the data, but rather its computation. It was this problem primarily which prompted the authors to develop the computer described in this paper.

A second factor influencing the development of a computer was the frequency with which simultaneous equations are encountered in engineering, in physics, and in mathematics; for example, such

equations occur in the study of electrical circuits, structures, aircraft flutter analysis, and statistics. Few computing devices have been available which rapidly yield solutions of engineering accuracy. The computer to be described will fill a need for such a device in many fields.

The present computer, in common with most previously proposed computers,¹⁻⁴ is of the analog type; i.e., certain physical quantities in the computer are used to represent, or are analogous to, the numbers involved. In the present computer the physical quantities involved are not voltage, current, or mechanical displacement, but are merely the *ratios* among certain resistances. This feature, together with a simple method of obtaining the summation of a number of quantities, contributes greatly to the simplicity and to the high inherent accuracy of the computer.

BASIC MATHEMATICAL PRINCIPLES

Before describing the details of the computer it is necessary to describe the mathematical method

upon which it is based. This method is known as the "Gauss-Seidel method" or the "Classical Iterative method."⁸ In it, an arbitrary solution, or an approximate solution if it is known, is used to compute a second approximation to the solution, which is then used to compute a third, and so forth until a negligible difference exists between successive approximations. These operations, although laborious by conventional methods, are extremely rapid in the present computer, as each step consists merely of adjusting a circuit to a null, with the instrument itself remembering the previous approximation as long as necessary.

For three equations the method is as follows: Let the set of equations be defined by

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = m_1, \quad (1)$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = m_2, \quad (2)$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = m_3. \quad (3)$$

In the following, a superscript on x (e.g., $x^{(p)}$) indicates the value existing at the end of the p th iteration.

First cycle:

a. The first equation is solved for $x_1^{(0)}$, using the trial values $x_2^{(0)}$ and $x_3^{(0)}$ which are arbitrary.

$$a_{11}x_1^{(1)} + a_{12}x_2^{(0)} + a_{13}x_3^{(0)} = m_1, \quad (4)$$

$$x_1^{(1)} = (1/a_{11})(m_1 - a_{12}x_2^{(0)} - a_{13}x_3^{(0)}).$$

b. The second equation is solved for $x_2^{(1)}$, using $x_1^{(1)}$ and $x_3^{(0)}$.

$$a_{21}x_1^{(1)} + a_{22}x_2^{(1)} + a_{23}x_3^{(0)} = m_2, \quad (5)$$

$$x_2^{(1)} = (1/a_{22})(m_2 - a_{21}x_1^{(1)} - a_{23}x_3^{(0)}).$$

c. The third equation is solved for $x_3^{(1)}$, using $x_1^{(1)}$ and $x_2^{(1)}$.

$$a_{31}x_1^{(1)} + a_{32}x_2^{(1)} + a_{33}x_3^{(1)} = m_3, \quad (6)$$

$$x_3^{(1)} = (1/a_{33})(m_3 - a_{31}x_1^{(1)} - a_{32}x_2^{(1)}).$$

Second cycle:

$$a'. \quad a_{11}x_1^{(2)'} + a_{12}x_2^{(1)} + a_{13}x_3^{(1)} = m_1. \quad (7)$$

$$b'. \quad a_{21}x_1^{(2)} + a_{22}x_2^{(2)'} + a_{23}x_3^{(1)} = m_2. \quad (8)$$

$$c'. \quad a_{31}x_1^{(2)} + a_{32}x_2^{(2)} + a_{33}x_3^{(2)'} = m_3. \quad (9)$$

The primed term is the one solved for in each equation. This procedure is repeated until the variables remain the same on two successive cycles. On the electric computer, this means that a null is obtained on switching to any equation.

For the general case, the method can be expressed concisely in matrix notation as $A_1X^{(p)} + A_2X^{(p-1)} = M_1$, where

$$A = \text{square matrix of the coefficients} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix},$$

$$A_1 = A \text{ with all elements above the principal diagonal equal to zero} = \begin{vmatrix} a_{11} & \cdots 0 & 0 \\ \cdot & & \cdot \\ \cdot & \cdots a_{ii} & 0 \\ \cdot & & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix},$$

$$A_2 = A \text{ with the principal diagonal and all elements below it equal to zero} = \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & a_{n-1,n} \\ 0 & \cdots & & 0 \end{vmatrix},$$

$$X = \text{column matrix of the unknowns} = \begin{vmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{vmatrix}, \quad M = \text{column matrix of the constant terms} = \begin{vmatrix} m_1 \\ \cdot \\ \cdot \\ \cdot \\ m_n \end{vmatrix}.$$

Convergence of the Method

Whether or not the iterative method converges depends upon the coefficients a_{ij} , and upon the order in which the equations and variables are placed before solution is begun. A necessary and sufficient condition for convergence of the method has been developed,⁵⁻⁸ but unfortunately it requires a large amount of computation to apply the criterion. More readily computed criteria may be deduced, but these are merely sufficiency criteria, and give no conclusive information.

It is, therefore, fortunate that in a great many practical problems a satisfactory order of arrangement resulting in convergence is obvious upon inspection. The simple empirical rules used in obtaining this order are:

1. An attempt is made to put the equations in such order that

- coefficients on the diagonal are large, and
- those off the diagonal—particularly the post-diagonal coefficients—are small.

When these rules can be applied, the problem usually converges in a few cycles.

2. In case rule (1) cannot be applied, a major coefficient in each equation is put on the diagonal. Then the order is adjusted until the product of each pair of diagonal elements is greater in absolute value than the product of the other two elements in the second-order determinant having those diagonal elements as the principal diagonal; i.e.,

$$|a_{11}a_{22}| > |a_{12}a_{21}|. \quad (10)$$

In this case, also, the smaller off-diagonal terms are placed in post-diagonal positions if possible.

Once a satisfactory order for a given set of coefficients is determined, that order can be used for all problems involving those coefficients, since the conditions for convergence (but not the *rate* of convergence) depend upon coefficients a_{ij} , only and not on constant terms m_i .

K-Method of Accelerating Solution

For cases in which convergence is slow and/or the optimum order is not immediately apparent, a direct method, called a *K*-method, has been developed. In this method it is only necessary to go through the process once to obtain the final answer; that is, there are no successive approximations to be made. Although the details are too involved to give here, briefly the method is as follows.* The first two equations in the first two

* It is hoped to publish the full details of this method at an early date.

variables are solved, then the first three equations in the first three variables, etc., until all equations in n variables have been solved. The coefficients of the successive sets of equations are those which each time include one more element of the principal diagonal of the matrix:

$$\begin{array}{cccc|cccc|cccc|cccc|} a_{11} & a_{12} & & & a_{13} & a_{14} & \cdots & a_{1n} & & & & & & & & \\ a_{21} & a_{22} & & & a_{23} & a_{24} & \cdots & a_{2n} & & & & & & & & \\ a_{31} & a_{32} & a_{33} & & a_{34} & & \cdots & a_{3n} & & & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} & & & & & & & & & & \end{array}$$

The method of accomplishing the successive solutions directly, without recourse to a trial and error procedure, is to balance each equation not to a null, but to a previously computed portion, k_i , of the unbalance observed when switching to that equation. The purpose of solving successively more equations 2, 3, $\cdots$ to n , as outlined above is to obtain k 's which are functions of the a 's and are independent of the m 's.

The process requires precomputation of the k 's but these can be readily calculated on the computer itself. There are $n(n-1)/2$ k 's to be computed for n equations. Since these are functions of the a 's only, time can sometimes be saved by utilizing this method when a number of solutions are to be made from the same coefficients but with different constant terms.

BASIC CIRCUITS

From Eqs. (4) through (9) it is evident that the two basic mathematical operations to be performed are multiplications of pairs of numbers, and additions or subtractions of the quantities which result.

Figure 1 shows the method for adding voltages which is used in the computer. In this, n equal impedances are connected to a common point O' at one end, and the other end of each is connected to one of the voltages to be added. The other side

of each voltage source is connected to another common point O , which may for convenience be ground. Then, applying Kirchoff's laws to the circuit, the open circuit voltage between O' and O is easily found to be

$$E_{oc} = (1/n)(E_1 + E_2 + E_3 + \cdots + E_n). \quad (11)$$

In other words, E_{oc} is proportional to the sum of the various source voltages. This is a general result, applying both to direct and alternating voltages, but it will be considered here only with regard to alternating voltages of the same phase and frequency, since the use of alternating current is found to have a number of practical advantages.

In order that an addition circuit of this type yield an accurate sum, two conditions must be satisfied: (a) the n impedances comprising the

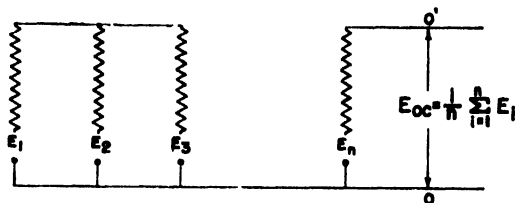


FIG. 1. Diagram of a circuit for obtaining a voltage proportional to the sum of a number of voltages.

circuit must be accurately equal in impedance, and (b), the internal impedances of the voltage sources must be negligibly small compared with the coupling impedances. Both of these conditions can be met satisfactorily in practice.

Figure 2 shows the circuit of a multiplying device, utilizing a pair of adjustable voltage dividers, or potentiometers, one supplying the other. If a voltage E_1 is connected across the first potentiometer and its slider is set to the fraction a of the total resistance (taking into account the loading effect of the second potentiometer), then the voltage applied to the second potentiometer is aE_1 . If the slider of the second potentiometer is set at the fraction x of the total resistance, then the output voltage is $E_0 = axE_1$. It should be noted that the second potentiometer, which may be called the x -potentiometer, is linear, whereas the a -potentiometer is non-linear because of its load. This feature is not particularly disadvantageous, because in the interest of the greatest possible accuracy it is desirable to set values on potentiometers not with reference to a calibrated

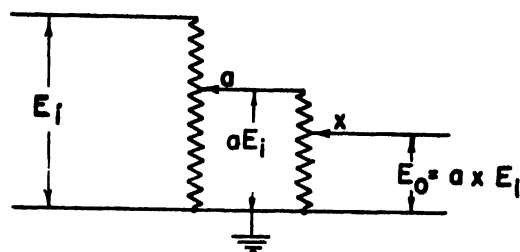


FIG. 2. Diagram of a circuit for obtaining a voltage proportional to the product of two numbers.

scale on the potentiometer itself, but rather with respect to a decade-type standard potentiometer.

In order to accomplish such comparative measurements, a simple Wheatstone bridge circuit is sufficient. Figure 3 shows a measuring circuit, in which the potentiometer to be set and an accurate decade potentiometer are connected in parallel to a common source. A null indicator of some sort is connected between their taps. It is plain that any fraction set on the decade can be transferred to the a -potentiometer by adjusting the a -potentiometer to a null, and that the loading effect of the x -potentiometer is thus automatically taken into account.

With these simple circuits as a basis it is easy to build an analogous circuit representing one equation of the type of (1). Figure 4 shows a combined multiplication-addition circuit which represents

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = m. \quad (12)$$

One half of a center-tapped transformer supplies a set of a -potentiometers in parallel, and between the tap and ground of each of these potentiometers is connected an x -potentiometer. Across the other half of the transformer is connected a single potentiometer which represents the m term. Recalling now the operation of a multiplying circuit, it is seen that the voltage between the arm of each x -potentiometer and ground has the

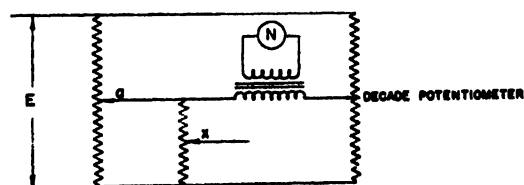


FIG. 3. Diagram of a circuit for setting a ratio on a loaded potentiometer by using a standard decade potentiometer.

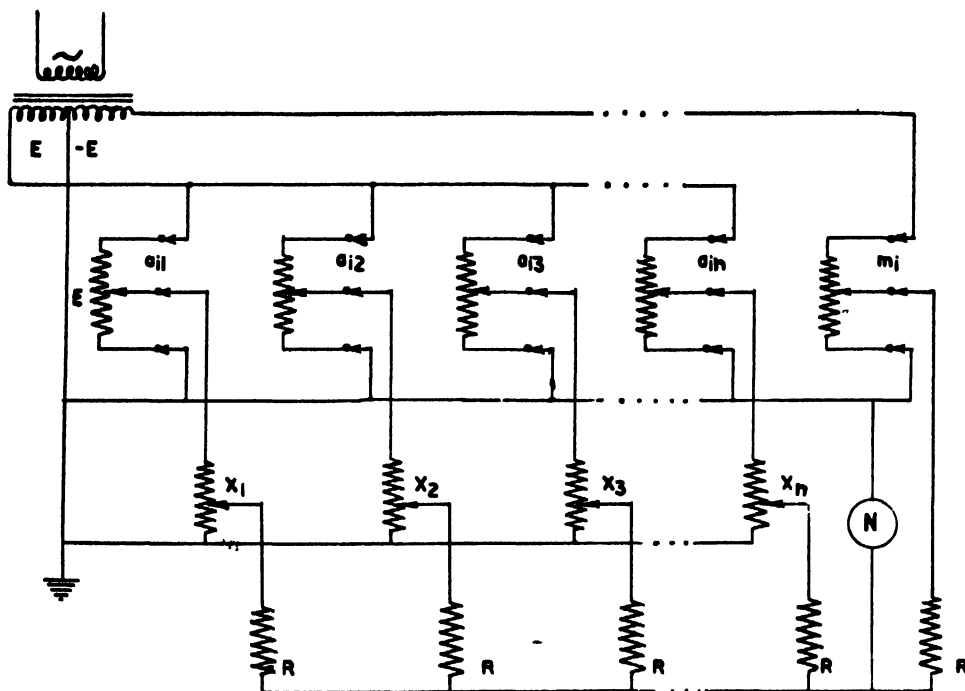


FIG. 4. Diagram of a circuit which is the analog of an equation of the type $\sum a_i x_i = m_i$.

value

$$V_{ax} = a_i x_i E, \quad (13)$$

and the voltage between the arm of the m -potentiometer and ground is simply

$$V_m = -m_i E, \quad (14)$$

where the minus sign must be used because this voltage is opposite in phase to that in Eq. (9). If the arms of the x - and m -potentiometers are regarded as voltage sources for a circuit of the type shown in Fig. 1, it is only necessary to connect them to a common point through equal high impedances in order to obtain a voltage proportional to the sum. This is accomplished in Fig. 4 by means of the resistors R . Thus the voltage between the junction point and ground in Fig. 4 is given by

$$V_n = [1/(n+1)](a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \dots + a_{in}x_n - m_i)(E). \quad (15)$$

Now, any adjustment of the x -potentiometers which results in a value of zero for V_n transforms Eq. (15) into Eq. (12), since the factors E and $1/(n+1)$ drop out. In other words, any set of x -potentiometer settings which causes V_n to be zero represents a solution for the particular equation in question.

The method of extending the circuit of Fig. 4 to sets of equations having any number of unknowns is obvious, since it is merely necessary to provide as many a - and x -potentiometers, each in combination with its associated addition resistor, as unknowns. Likewise, there must be provided as many rows of a - and m -potentiometers as there are equations, together with means for switching each row as a group (representing the coefficients and constant term of one equation) into the voltage source and the x -circuits as indicated in Fig. 4. For simplicity, only one set of a -potentiometers has been shown in Fig. 4.

The circuit shown in Fig. 4 is modified in the commercial computer in that two sets of addition resistors are used to permit positive and negative values to be handled; at the same time this eliminates the need for a center-tapped input transformer. The chief advantage of eliminating the transformer is that high accuracy is more easily obtained without the necessity of maintaining great precision of the center tap under varying load conditions on the two halves of the transformer. Figure 5 is a photograph of a commercial 12-equation computer designed according to the preceding theory.

The operation of a computer based on such a circuit is extremely simple. The coefficients and constant terms are first set in by use of a comparison bridge, as shown in Fig. 3. Then, with the a -potentiometers and m -potentiometer for the first equation switched into the circuit, the first x -potentiometer is adjusted to give a null; the second set of a 's and their associated m are switched into the circuit and the second x -potentiometer adjusted to satisfy the second equation, and so on. This sequence is repeated until all equations are found to be in balance. The x -potentiometers then read the values of the unknowns directly.

ACCURACY

The question of accuracy is of prime importance in determining the usefulness of any such computer as that described. In engineering appli-

cations one of the most useful criteria of accuracy is that the errors introduced by the computations (depending upon the problem) be either small compared with the errors inherent in the initial data, or at least be of the same magnitude as these errors. The initial data are normally the coefficients and constant terms of the equations. It has been found possible to design a computer based on the foregoing circuits such that, in using data accurate to between two and three significant figures, the errors in the x values obtained on the computer are about *one order of magnitude less* than the errors inherent in the data. In using three to four place data the accuracy of the computer is of the same order of magnitude as the accuracy inherent in the data.

A complete discussion of the subject of accuracy in the solution of linear simultaneous algebraic equations is beyond the scope of this

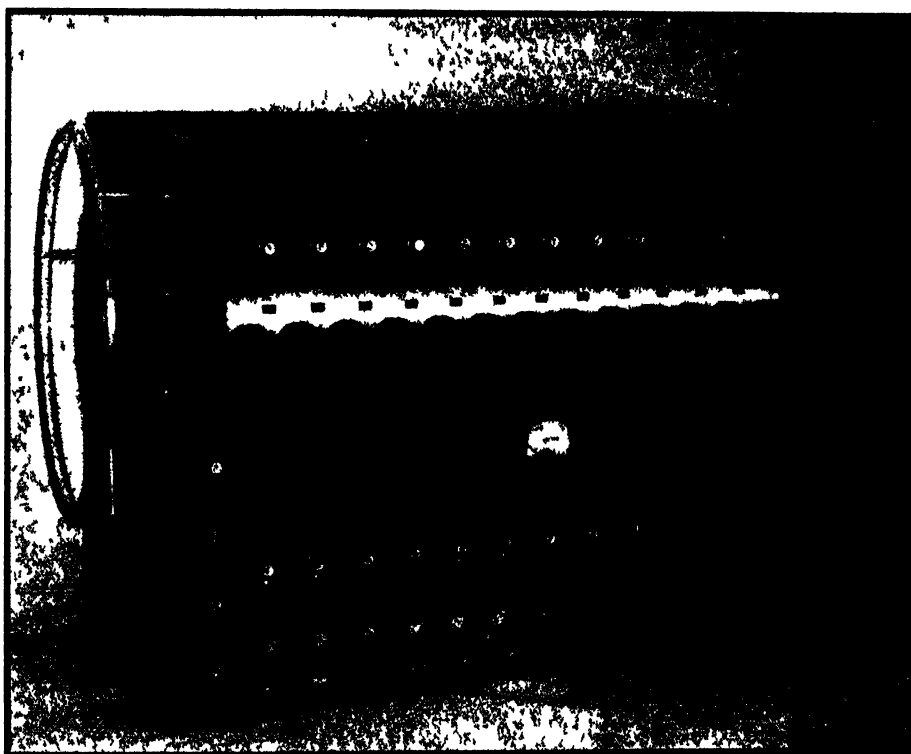


FIG. 5. Front view of a commercial 12-equation computer. The a - and m -potentiometers are carried on a switching turret operated by the large handwheel. The m 's and the a 's representing one equation appear in the horizontal window and are automatically switched into the circuit when turned to this position. The bottom row of dials represents the x values. The knobs above the x dials control plus-minus switches, each corresponding to the x dial below it. The vertical row of push-buttons at the right is for presetting positive and negative m 's. In the center of the panel is a small cathode-ray tube null indicator, and to the left of this are the four dials of the decade potentiometer. The length of the panel is 32 inches.

TABLE I. Accuracy and time of electric computer solution of twelve simultaneous algebraic equations. The first 12 rows and columns are the coefficients of the equations to be solved; each column representing the coefficients of one variable, each row the coefficients of one equation. The m column contains the constant terms; the m_{cc} column the results of substituting the electric computer answers in each equation.

Equation No.	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	m	m_{cc}	Δm = $m_{cc} - m$
(1)	1.0000	.1359	.1070	.0256	.0065	.0091	.0469	.0239	.0046	.0101	.0072	.0086	.1127	.1122	-.0005
(2)	.0278	1.0000	.1944	.1659	.1779	.0585	.2129	.1100	.0248	.0025	.0328	.0323	.1484	.1480	-.0004
(3)	.0020	.1573	1.0000	.0111	.1853	.1161	.2273	.1472	.0921	.1010	.0687	.0029	.1748	.1745	-.0003
(4)	.0015	.1920	.0318	1.0000	.7289	.0030	.0445	.0088	.0362	.0029	.0423	.0036	.1988	.1984	-.0004
(5)	.0050	.0346	.0391	.0822	1.0000	.0199	.0821	.0123	.0089	.0328	.0718	.0042	.1236	.1232	-.0004
(6)	.0216	.1759	.0429	.0254	.1242	1.0000	.1642	.0094	.0034	.0154	.0062	.0062	.0873	.0873	0
(7)	.0139	.1379	.0230	.0209	.0562	.2318	1.0000	.0176	.0028	.0056	.0044	.0103	.0668	.0668	0
(8)	.1834	.2222	.2103	.0739	.1938	.3345	.2938	1.0000	.0424	.0102	.0053	.0204	.2598	.2598	0
(9)	.0724	.0632	.0943	.1028	.0072	.0029	.0423	.4978	1.0000	.0018	.0028	.0183	.1791	.1791	0
(10)	.0474	.1947	.0610	.0725	.1880	.0608	.0222	.1024	.0621	1.0000	.0032	.0216	.0832	.0831	-.0001
(11)	.0683	.0457	.0241	.0759	.0139	.0173	.0034	.1521	.0583	.4523	1.0000	.0347	.1302	.1300	-.0002
(12)	.1124	.0320	.0238	.0970	.0908	.0104	.0085	.0383	.0326	.0364	.0121	1.0000	.2111	.2110	-.0001
Value of determinant															= .76978
Times															
Correct x 's	.0820	.0535	.1145	.1050	.0970	.0480	.0320	.1500	.0680	.0090	.0725	.1685	To set matrix (Prorated over 10 solutions)		3.6 min.
Electric computer x 's	.0816	.0532	.1143	.1050	.0966	.0481	.0320	.1502	.0680	.0090	.0723	.1685	To set m 's		3.0 min.
													To balance		2.0 min.
													To read x 's		3.0 min.
													Total average time per solution		11.6 min.
Δx	-.0004	-.0003	-.0002	0	-.0004	.0001	0	.0002	0	0	-.0002	0	Time by conventional calculator method	80	min.

paper. In general, however, each new set of coefficients must be separately investigated.^{5,6,9} Some of the unknowns in a single problem may be accurately determined, while others may be very sensitive to small variations in either a 's or m 's. In routine calculation on a conventional calculator such as the Marchant, Friden, or Monroe, it is often necessary to carry a large number of decimal places^{5,9} during the solution of simultaneous equations to obtain three or four figure accuracy in the answers, assuming the coefficients subject to no error. If coefficients are accurate to only four decimal places the unknowns may be accurate to only two or three decimal places, *regardless of how many places* are carried in intermediate steps. Generally, if upon substitution of the values of the unknowns back in the original equations, the resulting constant terms differ from the actual constant terms used by amounts comparable to the possible uncertainties in those constant terms, then the errors in the unknowns are those inherent in the data and no improvement can be made by any method of computation.

The design of the commercial computer shown in Fig. 5 is such that numbers from 0.0000 to 0.1000 can be set to about ± 0.0001 , and numbers from 0.1000 to 1.0000 can be set to one part in a thousand (or better) of the value being set. The values of the x 's can be read with the same

accuracy; however, the absolute accuracy of the x 's, as pointed out before, depends in large measure upon the particular set of equations in question.

There is a close connection between the inherent accuracy of solution and the nearness to singularity of the set of equations. When coefficients of two equations or of two variables in a set of n equations are proportional, the system has a zero determinant and is said to be singular. Two of the variables or relations are then dependent, and a unique solution for the n variables cannot be obtained. If the coefficients in two rows or two columns of the determinant are nearly proportional the system is nearly singular, and some of the x values will be very sensitive to changes in certain of the a 's and m 's. In this case more cycles of iteration will be required in the solution and the system may be said to have poor "leverage." The size of the determinant will usually be an index of this characteristic when the a 's are positive. For example, consider two sets of equations:

$$\begin{array}{ll} \text{A} & \text{B} \\ 1.00x_1 + 0.10x_2 = 0.50, & 1.00x_1 + 0.90x_2 = 0.50, \\ 0.10x_1 + 1.00x_2 = 0.50, & 0.90x_1 + 1.00x_2 = 0.50, \\ \text{Determinant} = 0.99; & \text{Determinant} = 0.19. \end{array}$$

The leverage of system A is better than that of system B.

TABLE II. Accuracy and time of electric computer solution of equations used to analyze a C₂ through C₄ paraffin-olefin mixture from mass spectrometer data.

Mass	Pen- tenes	Iso- pentane	N-pen- tane	Pro- pane	Ethane	N-bu- tane	Iso- butane	Ethylene	Pro- pene	Butene-1	Iso- butene	Butene-2	<i>m</i>	<i>m_{ee}</i>	Δm
70	1.0000	.0021	.0032	0	0	0	0	0	0	0	0	0	.0090	.0088	-.0002
57	.0017	1.0000	.3595	0	0	.0580	.0912	0	0	.0374	.0440	.0444	.0476	.0476	0
72	0	.4727	1.0000	0	0	0	0	0	0	0	0	0	.0098	.0096	-.0002
44	0	.1033	.1619	1.0000	0	.1437	.1774	0	.0072	.0020	.0006	.0017	.1524	.1524	0
30	.0355	.0316	.0274	.0810	1.0000	.0422	.0067	.0051	0	.0233	.0233	.0281	.0408	.0408	0
58	0	.1412	.0500	0	0	1.0000	.2966	0	0	.0027	.0016	.0026	.1394	.1394	0
43	.0134	.5920	.9255	.1507	0	.8076	1.0000	0	.0127	.0008	.0013	.0009	.3122	.3123	.0001
26	.1446	.0702	.0903	.1917	.5226	.1498	.0667	1.0000	.1770	.1089	.2135	.2365	.1758	.1756	-.0002
42	.7143	.9620	1.0000	.0708	0	.1836	.6274	0	.6566	.0415	.0440	.0369	.3590	.3587	-.0003
41	.4143	.6010	.5525	.1198	0	.3299	.5466	0	.6763	1.0000	.9845	.9007	.6218	.6215	-.0003
56	.1027	.3069	.0618	0	0	.0168	.0101	0	0	.8214	.9774	1.0000	.2889	.2890	.0001
55	1.0000	.0337	.0301	0	0	.0090	.0051	0	0	.1403	.1275	.1594	.0541	.0539	-.0002
Value of determinant = .00306															
<i>x</i>	.0090	.0080	.0060	.0980	.0200	.0700	.2270	.0150	.2510	.0680	.0950	.1330	To set matrix (prorated over 10 solutions) 3.6 min.		
<i>m_{ee}</i>	.0088	.0080	.0058	.0980	.0200	.0700	.2273	.0148	.2507	.0672	.0956	.1332	To set <i>m</i> 's 3.0 min.		
													To balance 9.0 min.		
													To read <i>x</i> 's 3.0 min.		
Δx	-.0002	0	-.0002	0	0	0	.0003	-.0002	-.0003	-.0008	.0006	.0002	Total average time per solution 18.6 min.		
													Time by conventional calculator method 80 min.		

In the examples which follow, the value of the determinant is given as an approximate index of the leverage of each system. Usually, the larger determinants are associated with better leverage and hence with better accuracy and a shorter time for solution.

It is interesting to note that the computer can be used to compute reciprocal matrices or inverses, using the method described by Fisher.¹⁰ The Fisher method involves solving *n* sets of *n* equations each to obtain the *n* columns of the inverse. Here again the intrinsic "leverage" is a substantial factor in determining accuracy. Where greater accuracy is desired in the inverse, the matrix obtained on the electric computer can be used as an initial approximation. Then

iterative methods employing either a calculating machine or combining the calculating machine and electric computer can be used to improve the accuracy.

In order to explore the possibilities of the computer alone without auxiliary computation a large number of problems have been solved on a 12-equation model. The results shown in Tables I-IV illustrate the accuracy which has been obtained in solving various typical problems. In Tables I-III, the coefficients, *a_{ij}*, and the constant terms, *m_i*, are tabulated. The three rows beneath this matrix show, respectively, the answers obtained on an 8-bank calculator, the electric computer answers, and the difference between the two, Δx . Two additional columns are

TABLE III. Accuracy and time of electric computer solution of equations involving both positive and negative coefficients and constant terms.

Equation No.	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	m	m_{ee}	Δm
(1)	1.0000	-.0811	.0414	.0018	.0090	0	0	-.0476	.0925	.0925	0
(2)	-.1290	1.0000	.5280	-.4650	.4170	.2910	0	.0342	-.1654	-.1654	0
(3)	0	-.2980	1.0000	0	0	0	0	0	.0927	.0927	0
(4)	.0106	.0207	-.0312	1.0000	.7100	-.6650	0	.0012	.0777	.0779	.0002
(5)	.0035	.0270	.0082	.8850	1.0000	-.6400	0	.0004	.0884	.0885	.0001
(6)	.0700	-.0631	.0854	.9650	-.8850	1.0000	.4530	.0227	.0741	.0738	-.0003
(7)	-.0870	.2170	.3350	-.0958	.1005	.0784	1.0000	-.0532	.1505	.1505	0
(8)	.2380	.3630	-.8390	.0050	.0044	.0021	.0440	1.0000	.3041	.3041	0
									Value of determinant		=.59314
x	.0930	-.1870	.0370	.0258	.0518	-.0282	.2060	.3715	To set matrix (prorated over 10 solutions)		1.6 min.
									To set m 's		2.0 min.
x_{ee}	.0930	-.1869	.0370	.0258	.0516	-.0286	.2060	.3715	To balance		8.0 min.
									To read x 's		2.0 min.
Δx	0	.0001	0	0	-.0002	.0004	0	0	Total average time per solution		13.6 min.

TABLE IV. Accuracy and time for solution of equations in spectrometric analysis of typical hydrocarbon mixtures. The components are different for the different columns, ranging from preponderantly butanes and butenes to all pentanes and hexanes.

Infra-red C ₄	C ₄ , C ₅	C ₃ -C ₄	C ₂ -C ₃	C ₄ -C ₅	C ₂ -C ₃ , CP	C ₄ , C ₅
0	0	.0002	0	-.0001	.0032	-.0003
.0002	-.0005	0	-.0001	0	-.0042	-.0057
-.0003	.0003	.0003	.0001	0	.0002	-.0056
.0001	.0002	.0019	-.0005	.0002	-.0005	-.0058
.0001	-.0003	-.0020	.0002	.0001	0	.0179
.0004	-.0003	.0001	.0007	.0010	.0013	-.0005
.0001	-.0005	-.0001	-.0010	-.0003	.0033	-.0001
.0002	.0002	0	-.0003	-.0009	-.0039	.0003
Max. error	.0004	.0005	.0020	.0010	.0042	.0179
Mean error	.0002	.0003	.0006	.0004	.0021	.0045
Value of determinant	.78908	.05931	.02891	.02221	.00590	.00080
Time to balance (min.)	4	8	8	10	11	13
Time to balance using <i>k</i> -method (after computation of <i>k</i> 's)				10 min.		20

given. The m_{ec} column is the result of substituting the electric computer answers into the original equations, using an 8-bank calculator. The Δm column shows the difference ($m - m_{ec}$) between these values and the actual constant term column used.

Table I illustrates a typical 12-equation system with large diagonal terms and a fairly large determinant. This set required only 5 cycles (2 minutes) to converge. The accuracy, as would be expected, is good, all errors in x 's being 0.0004 or less, compared with a total of 1.0000 for the sum of all the x 's.

Table II shows a mass spectrometer analysis for a 12-component hydrocarbon mixture including all paraffins from ethane through *iso*-pentane and a number of olefins through C₆. Coefficients in the last three columns are nearly proportional, so that errors in the last three variables are larger than in the others. Even so, the errors in these variables are in the fourth decimal place. The largest error is only 0.08 mole percent.

The fact that the m 's check to 0.0003 or better indicates that the electric computer errors are equivalent to errors of this magnitude in the original data. Thus again the x errors are of the same order of magnitude as those which would be expected in a physical problem if the m data were subject to error in the fourth place.

A problem similar to those encountered in aircraft structural analysis, involving some nega-

tive constant terms and some negative a 's, is given in Table III. Positive or negative m 's can be set, and positive or negative x 's read. Means of pre-setting positive and negative a 's can be built into the computer if desired, but with the present computer a change of sign in an a_i value is accomplished by changing the sign of its associated x on switching to a new equation. Otherwise the method of balancing is the same as in the case of positive coefficients.

Table IV summarizes the results of solving seven different sets of eight equations each. In each of these problems a hydrocarbon mixture is analyzed, the first from infra-red data and the remaining 6 from mass spectrometer data. Hence all a 's are positive. The error in each component, the maximum error, the average error, and the time for solution are shown. The value of the determinant is given as a general leverage index. As would be anticipated from the fact that different variables are widely different in their response to changes in the coefficients, accuracies in the values of some unknowns are much higher than in others. However, the time of solution and the average error decrease fairly systematically as the determinant increases, as would be expected.

The computer is not, of course, limited to solving sets of 12 equations. Any number up to 12 may also be handled.

SPEED

The speed of operation of a computer chiefly determines its advantages over conventional

methods of calculation. In the computer described a large saving of time is effected, which becomes more and more apparent as the number of equations to be solved increases. For example, a reasonably expert calculator would require nearly 5 hours (assuming no errors) to solve by conventional machine methods the set of 12 equations given in Table I, as compared with 44 minutes for the first solution on the electric computer. In other words, only 1/7 as much time would be required when using the computer.

In Tables I-III, times of solution are given in which the time for setting the coefficients has been prorated over ten different solutions. The reason for this is that in many applications (e.g., hydrocarbon mixture analysis) a number of mixtures involving the same components are solved successively. Hence the same coefficients are used a number of times with different constant terms. To obtain a realistic picture of the times involved some such weighting is therefore necessary. The comparable average figure for solving 10 sets of 12 equations each involving the same coefficients on a calculator is about 80 minutes per set. The computer times are only $\frac{1}{4}$ to $\frac{1}{7}$ the times required by a calculator.

The times given in Table IV are only those required for balancing and are nearly proportional to the number of iterative cycles. It is interesting to note the relation illustrated here between the value of the determinant and the time for balancing; as the determinant increases the time decreases.

One of the primary factors contributing to the speed of the computer is the fact that each of the coefficients is handled only once—when it is originally set on its corresponding potentiometer. This is in contrast to conventional methods in which the operator has to set many more than n^2+n numbers on the keyboard and also has to write down intermediate results. Eliminating these intermediate steps eliminates also one of the most common sources of error in conventional methods of solution.

In addition a relatively unskilled operator usually finds it simpler to adjust dials to obtain a null than to operate a calculating machine, so that facility can be acquired in a very short time. If the wrong dial is turned during the balancing

operation, the iterative method compensates for the mistake. Although a few extra cycles may be necessary in this case, the final answer is still correct. So long as the a 's and m 's are initially set correctly, nulls on all equations indicate that the correct solution has been obtained, regardless of the intermediate steps taken. Thus not only the probability of errors, but also their adverse consequences are reduced to a minimum.

Further time saving results from the fact that each operation is brief. To set each coefficient, a_{ij} , and each constant term, m_i , requires about $\frac{1}{4}$ minute. The time to balance the equations by successive iterative cycles varies with the intrinsic leverage of the equations and also with the order of equations and unknowns which is chosen for balancing. One cycle for nine equations requires about one minute. The time to read each unknown is less than $\frac{1}{4}$ minute.

In using the K -method the time required to solve a given set of equations depends only upon the number of equations, since the method is direct rather than iterative.

To compute K 's for eight equations takes about 38 minutes, after which use of the K 's in the solution requires about 10 minutes. This figure can be compared with the times shown in Table IV for balancing by the iterative method. The K -method is particularly useful in those cases in which convergence is slow and where the order in which variables and equations should be placed is not immediately evident. Usually, if the set of equations can be arranged to converge fairly rapidly, the iterative method is more rapid than the K -method.

CONCLUSION

The computer described may be applied in almost any field in which simultaneous linear equations must be solved. The fact that it can be used on either symmetric or non-symmetric matrices extends its application to many engineering fields. For instance, some problems arising in aircraft structural design have been solved.

In the field of statistics, it is believed that it will prove highly useful not only in solving equations, but also in providing an initial approximation to reciprocal matrices.

It is now being used with eminent success in

the field of hydrocarbon analysis with the mass spectrometer and infra-red spectrophotometer. Here, its accuracy is more than adequate, and its speed is far greater (in many cases by a factor of 5 or more) than that of conventional methods.

The authors wish to acknowledge the fine work of Mr. J. C. Pemberton, who engineered the commercial model of the computer, and of Miss Dorothy Comaford, who accumulated the data given in the tables. The authors also wish to thank Consolidated Engineering Corporation for permission to publish this material.

BIBLIOGRAPHY

- (1) C. H. Ward, U. S. Patent 1,893,009, Jan. 3, 1944.
- (2) R. R. M. Mallock, "Electrical calculating machine," *Proc. Roy. Soc. London* **140**, 457-482 (1933).
- (3) O. Wilbur, "The mechanical solution of simultaneous equations," *J. Frank. Inst.* **222**, 715-724 (1936).
- (4) T. D. Morgan and F. W. Crawford, "Time saving computing instruments designed for spectroscopic analysis," *Oil Gas J.*, pp. 100-105 (August 26, 1944).
- (5) H. Hotelling, "Some new methods of matrix calculation," *Ann. Math. Stat.* **14**, 1-34 (1943).
- (6) F. E. Satterthwaite, "Error control in matrix calculation," *Ann. Math. Stat.* **15**, No. 4, 373-387 (1944).
- (7) R. von Mises and Hilda Pollaczek-Geiringer, "Zusammenfassende Berichte. Praktische Verfahren der Gleichungsauflosung," *Zeits. f. angew. Math. Mechanik* **9**, 58-77, 152-164 (1929).
- (8) C. E. Berry, "A criterion of convergence for the classical iterative method of solving linear simultaneous equations," *Ann. Math. Stat.* **16**, 398-400 (1945).
- (9) L. A. Tuckerman, "On mathematically significant figures in the solution of simultaneous linear equations," *Ann. Math. Stat.* **12**, 307 (1941).
- (10) R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Ltd., London, 1938), seventh edition.

Other pertinent references not mentioned in the text are:

- A. C. Aitken, "Evaluation of determinants, formation of their adjugates, and practical solution of simultaneous linear equations," *Proc. Edin. Math. Soc.* [2] **3**, 207 (1932-33).
- A. C. Aitken, "Studies in practical mathematics," *Proc. Roy. Soc. Edin.* **57**, 269 (1937).
- A. T. Lonseth, "Systems of linear Eq. with coeff. subject to error," *Ann. Math. Stat.* **13**, 332-337 (1942).
- R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices and Some Applications to Dynamics and Differential Equations* (Cambridge University Press, London, 1938).

The Thin Cylindrical Antenna: A Comparison of Theories

DAVID MIDDLETON AND RONOLD KING

Cruft Laboratory and the Research Laboratory of Physics, Harvard University, Cambridge 38, Massachusetts

(Received November 17, 1945)

The solutions of Hallén's integral equation for the cylindrical antenna given by Hallén, Bouwkamp, Gray, King and Middleton are compared. Comparison is also made with the transmission-line theory of Schelkunoff and with experiment.

I. INTRODUCTION

BECAUSE of its great practical importance the antenna has been the subject of much analysis and discussion. In particular, considerable effort has been devoted to the thin, cylindrical antenna, and much has been written in the attempt to bring theory and experiment into closer agreement, and to give reliable theoretical results for such important and difficult-to-measure quantities as the input impedance and the distribution of current. Summaries are given by Brillouin¹ and by R. King and D. D. King.²

To determine the input impedance one must first establish the distribution of current. This is rigorously a three-dimensional problem, but in the case of thin antennas, where the thickness is a negligible fraction of a wave-length, a one-dimensional analysis suffices. A powerful approach has been suggested by Hallén,³ who obtains an integral equation involving the current as a solution of the one-dimensional wave equation for the vector potential. Recently King and Blake⁴ and King and Harrison,⁵ using Hallén's method, have given detailed computations for the input impedance and current distribution. Since agreement with the limited available experimental data was not particularly good, a more careful study of the experimental problem,^{2,6} and a more critical examination of theory^{1,2,7} was begun. Hallén's integral-equation method appears to be physically sound for thin antennas.

The difficulties lie in the method of solution and in the assumptions made in going from the actual physical problem to the simplified one of the analysis.² A modified form of solution has been suggested by Gray,⁸ to bring the theoretical values of impedance more in line with the limited existing experimental ones, and with the theoretical values obtained by S. A. Schelkunoff,⁹ who employs a different approach to the problem. More recently, the authors have achieved a solution¹⁰ which, it is believed, is superior to previous attempts.

It is the purpose, then, of this paper to compare the theories developed by (1) Hallén-Bouwkamp,¹¹ (2) Gray,⁸ and (3) King-Middleton,¹⁰ and to discuss the various solutions from theoretical and experimental points of view.^{2,12} A brief outline of the principal analyses follows.

II. ANALYSIS

The fundamental integral equation for the current, obtained originally by Hallén and containing the vector potential A_z , is

$$\int_{-h}^h I_z' \frac{\exp(-j\beta R_1)}{R_1} dz' = \frac{-4\pi j}{R_c} \left[C_1 \cos \beta z + \frac{1}{2} V_0 \sin \beta |z| - z' \int_0^s I(s) \sin \beta(z-s) ds \right] = \frac{4\pi A_z}{\Pi}, \quad (1)$$

with

$$R_1 = ((z-z')^2 + a^2)^{1/2}, \quad \beta = 2\pi/\lambda = \frac{2\pi f}{c}, \quad (2)$$

¹ L. Brillouin, *Elect. Comm.* 21, No. 4, 22, No. 1.

² R. King and D. D. King, *J. App. Phys.* 16, 445 (1945).

³ E. Hallén, *Nova Acta, Roy. Soc. Uppsala* 11, 1 (1938); see also C. W. Oseen, *Arkiv. f. Mat. Astr. Fysik* 9, No. 12, No. 30 (1913).

⁴ R. King and F. G. Blake, *Proc. I.R.E.* 30, 335 (1942).

⁵ R. King and C. W. Harrison, Jr., *Proc. I.R.E.* 31, 548 (1943).

⁶ D. D. King, *J. App. Phys.* 16, 435 (1945).

⁷ R. King and C. W. Harrison, Jr., *J. App. Phys.* 15, 170 (1944).

⁸ M. C. Gray, *J. App. Phys.* 15, 61 (1944).

⁹ S. A. Schelkunoff, *Proc. I.R.E.* 29, 493 (1941); *J. App. Phys.* 15, 54 (1944).

¹⁰ R. King and D. Middleton, *Quart. App. Math.* 3, 302 (1946).

¹¹ C. J. Bouwkamp, *Physica* 9, 609 (1942).

¹² G. H. Brown and O. M. Woodward, *Proc. I.R.E.* 33, 257 (1945).

and where V_0^e , R_c and z^i are, respectively, the applied potential difference maintained by a "slice" generator² at $z=0$, a characteristic resistance $R_c = 120\pi = 376.7$ ohms, and the internal impedance per unit length of the antenna. For good conductors z^i is sufficiently small so that the term involving z^i may be dropped in (1) with negligible error.⁴ C_1 is a constant of integration determined from the boundary condition that requires the current to be zero at the ends $z = \pm h$. Π is $4\pi \cdot 10^{-7}$ henry/m, and a is the radius of the antenna. The general m th order solution for the currents in (1) may then be obtained in the form of a series^{3-5, 7, 8, 10, 11}

$$(I_z)_m = \frac{j2\pi V_0^e}{R_c \Psi} \left\{ \frac{\sin \beta(h - |z|) + \sum_{n=1}^m M_n(z)/\Psi^n}{\cos \beta h + \sum_{n=1}^m F_n(h)/\Psi^n} \right\},$$

$$-h \leq z \leq h, \quad h^2 \gg a^2, \quad (3)$$

where Ψ is an appropriately chosen expansion parameter, defined by

$$\Psi \equiv |\Psi(z_0)|, \quad (4)$$

with z_0 a suitably selected point on the antenna. The function $\Psi(z)$ may be obtained from

$$\Psi(z) = \int_{-h}^h g(z, z') \frac{\exp(-j\beta R_1)}{R_1} dz' \equiv \Psi + \gamma(z), \quad (5)$$

and $\gamma(z)$ is defined by

$$\gamma(z) \equiv \Psi[\exp(j\beta\psi) - 1]. \quad (6)$$

Here $g(z, z')$ is a quantity whose character depends on the assumptions made for the distribution of current along the antenna. If $\Psi(z)$ is predominantly real, $\gamma(z)$ is a small complex correction function except at values of z where I_z is small or zero. It is to be noted that $\gamma(z)$ is infinite at $z = \pm h$ but that $I_z \gamma(z)$ is small and finite there. It is also possible that $\Psi(z)$ be entirely real; (see Eq. (17) following). Under these circumstances one may still write (5), but (6) no longer applies; $\gamma(z)$ may then be considered as a small, *real* correction factor, except at points where I_z approaches zero.

By definition $M_n(z)$ is the coefficient of Ψ^{-n} in

$$\sum_{n=0}^m F_{nz}/\Psi^n \cdot \sum_{n=0}^m G_n(h)/\Psi^n - \sum_{n=0}^m G_{nz}/\Psi^n \cdot \sum_{n=0}^m F_n(h)/\Psi^n, \quad (7)$$

where in particular

$$M_1(z) \equiv M_1^I(z) + jM_1^{II}(z) = F_1(z) \sin \beta h - F_1(h) \sin \beta |z| + G_1(h) \cos \beta z - G_1(z) \cos \beta h, \quad (8)$$

$$M_2(z) \equiv M_2^I(z) + jM_2^{II}(z) = F_2(z) \sin \beta h - F_2(h) \sin \beta |z| + G_1(h) F_1(z) - G_1(z) F_1(h) + G_2(h) \cos \beta z - G_2(z) \cos \beta h. \quad (9)$$

The F - and G -functions satisfy the following:

$$F_n(z) = F_{n-1,z} \int_{-h}^h g(z, z') \frac{\exp(-j\beta R_1)}{R_1} dz' - \int_{-h}^h F_{n-1,z'} \frac{\exp(-j\beta R_1)}{R_1} dz' - F_{n-1,z} \gamma(z), \quad (10)$$

with similar expressions for $G_n(z)$ obtained on replacing F by G in (10) above. The quantities F_{nz} , G_{nz} are defined, respectively, as $F_{nz} \equiv F_n(z) - F_n(h)$, $G_{nz} \equiv G_n(z) - G_n(h)$, from which it is clear that $F_{nh} = G_{nh} = 0$. When $n=0$ one has

$$F_0(z) = \cos \beta z, \quad G_0(z) = \sin \beta |z| \quad (11)$$

for all expansions. The input impedance follows at once from the definition

$$(Z_0)_m \equiv V_0^e / (I_z)_m = (R_0)_m + j(X_0)_m. \quad (12)$$

(a) Hallén³-Bouwkamp¹¹ solution

The relative distribution function for the current is taken to be

$$g(z, z')_H = \exp(j\beta R_1), \quad (13)$$

on the assumption that the vector potential A_z in (1), may be evaluated approximately by considering the current at all points to be I_z and neglecting the effects of retardation. The contributions from all points $z' \neq z$ contribute little, so that A_z depends primarily on the current at or near z . For this distribution (5) becomes

$$\Psi(z)_H = \sinh^{-1} \left(\frac{h-z}{a} \right) + \sinh^{-1} \left(\frac{h+z}{a} \right), \quad (14)$$

and in particular with $h^2 \gg a^2$

$$\Psi(0)_H = 2 \log \frac{2h}{a} \equiv \Omega. \quad (15)$$

Here the expansion parameter is chosen to be

$$\Psi = \Psi(0)_H = \Omega, \quad z_0 = 0, \quad (16)$$

from (4); $\gamma(z)_H$, a small real quantity, significant only within a few radii a of the ends $z = \pm h$, is

$$\gamma(z)_H = \log \left(1 - \frac{z^2}{h^2} \right) + \delta(z), \quad (17)$$

where

$$\delta(z) = \log \left[\frac{1}{4} \left\{ 1 + \left(1 + \frac{a^2}{(h+z)^2} \right)^{\frac{1}{2}} \right\} \times \left\{ 1 + \left(1 + \left(\frac{a}{h-z} \right)^2 \right)^{\frac{1}{2}} \right\} \right]. \quad (18)$$

(The subscript H is used throughout this paper to refer to the Hallén-Bouwkamp solution.) Eq. (3) results for the m th order current distribution, with Ψ replaced by Ω and the subscripts H written on $M_n(z)$ and $F_n(h)$. Fairly extensive tables of these functions are available^{4,10,11} for $n=1, n=2$, where

$$\begin{aligned} F_n(h) &\equiv \alpha_n \equiv \alpha_n^I + j\alpha_n^{II}, \\ \text{and} \quad M_n(0)^I + jM_n(0)^{II} &\equiv \beta_n = \beta_n^I + j\beta_n^{II}, \end{aligned} \quad (19)$$

and where general relations for $F_n(z)_H$ are given by

$$\begin{aligned} F_n(z)_H &= (F_{n-1,z})_H \Omega \\ &- \int_{-h}^h (F_{n-1,z'})_H \frac{\exp(-j\beta R_1)}{R_1} dz', \end{aligned} \quad (20)$$

with similar expressions for $G_n(z)_H$, obtained as before by replacing F by G in (20).

The input impedance is at once seen to be

$$(Z_0)_{mH} = \frac{-j\Omega R_c}{2\pi} \left\{ \frac{\cos \beta h + \sum_{n=1}^m \alpha_n / \Omega^n}{\sin \beta h + \sum_{n=1}^m \beta_n / \Omega^n} \right\}. \quad (21)$$

Hallén's work included the zeroth and first approximations, ($n=0, 1$), while Bouwkamp extended the solution through the second-order terms ($n=0, 1, 2$). Curves for the first- and second-order impedances computed from (21) are shown in Figs. 1-4 for $\Omega=10, 20$.

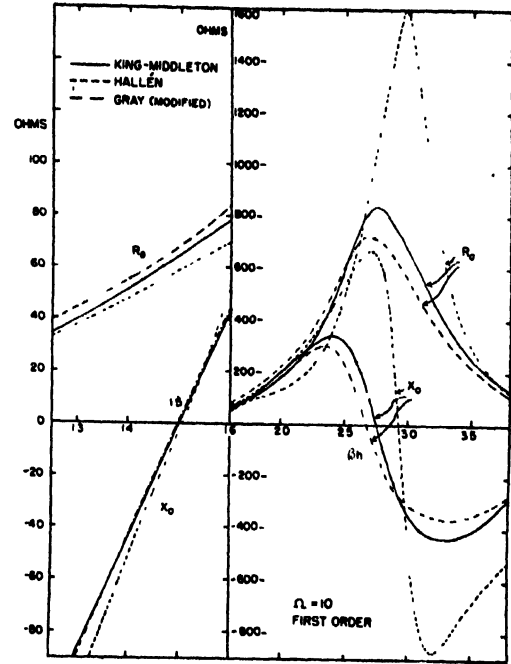


FIG. 1. First-order impedances $Z_0 = R_0 + jX_0$ of cylindrical antenna with $\Omega = 2 \ln 2h/a = 10$.

We give now, in somewhat greater detail as the results are new, a solution based on a distribution function suggested by, but differing in important respects from one introduced by, Gray.⁸

(b) Modified Gray Solution

In this case g is chosen to be

$$g(z, z')_G = \exp(j\beta R_1) \cos \beta R_1, \quad (22)$$

so that $\Psi(z)_G$ becomes, with the help of (5),

$$\Psi(z)_G = \Psi(z)_H - \text{Cin } \beta(h+z) - \text{Cin } \beta(h-z), \quad (23)$$

where, as usual,

$$\text{Cin } x \equiv \int_0^x \frac{1 - \cos q}{q} dq, \quad \text{Si } x \equiv \int_0^x \frac{\sin q}{q} dq. \quad (24)$$

(These and associated functions are considered in detail in the appendix of reference 10.) Instead of setting $z=0$ as does Hallén, we follow Gray for the moment and use the *average* value of $\Psi(z)_G$ as expansion parameter, *viz.*

$$\Psi_G = \langle \Psi(z)_G \rangle_m = \frac{1}{2h} \int_{-h}^h \Psi(z)_G dz, \quad (25)$$

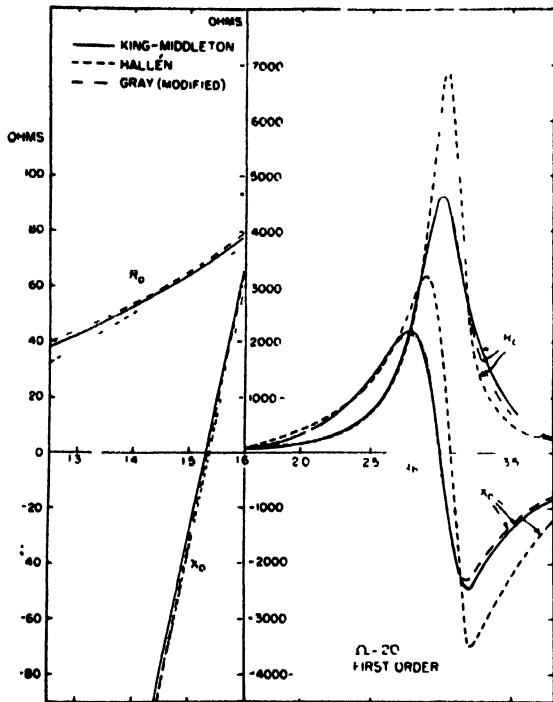


FIG. 2. First-order impedances; $\Omega = 20$.

which integrates to give

$$\Psi_G \equiv \bar{\Omega}_G = \Omega - 2K, \quad (26)$$

where

$$K \equiv \text{Ci} \, 2\beta h + \frac{\sin 2\beta h}{2\beta h} - \log 2; \quad \gamma(z)_G = 0. \quad (27)$$

Curves showing $\bar{\Omega}_G$ as a function of βh , with Ω as parameter, are given in Fig. 5. As in the Hallén solution the value of the current at $z = \pm h$ is subtracted from the fundamental integral Eq. (1), in order that the complete integral solution may vanish for *all* values of βh at $z = \pm h$. (See the paragraph following Eq. (18), reference 10.) When this is done, (10) and (22) give the general result

$$F_n(z)_G = (F_{n-1,z})_G \bar{\Omega}_G - \int_{-h}^h (F_{n-1,z'})_G \frac{\exp(-j\beta R_1)}{R_1} dz', \quad (28)$$

analogous to (20). Since $F_0(z)_H = F_0(z)_G$, etc., (28) yields at once

$$\begin{aligned} F_1(z)_G &= F_1(z)_H - 2K(F_{0z})_H; \\ F_1(h)_G &= F_1(h)_H; \\ (F_{1z})_G &= (F_{1z})_H - 2K(F_{0z})_H, \end{aligned} \quad (29)$$

and from (29) in (28) there follows for $n=2$

$$(F_{2z})_G = (F_{2z})_H + 2(-2K)(F_{1z})_H + (-2K)^2(F_{0z})_H; \quad (30)$$

$$F_2(h)_G = F_2(h)_H + (-2K)F_1(h)_H.$$

This process may be repeated to give

$$\left(\frac{F_n}{G_n} \right)_G = \sum_{i=0}^n \frac{n!}{(n-i)!} (\bar{\Omega}_G - \Omega)^i \left(\frac{F_{n-i}}{G_{n-i}} \right)_H, \quad n \geq 0, \quad (31)$$

and

$$\left(\frac{F_n(h)}{G_n(h)} \right)_G = \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-1-i)!} (\bar{\Omega}_G - \Omega)^i \times \left(\frac{F_{n-i}(h)}{G_{n-i}(h)} \right)_H, \quad n \geq 0, \quad (32)$$

since $-2K = \bar{\Omega}_G - \Omega$, cf. Eq. (26). The functions $M_n(z)_G$ formed from F and G in the manner of (7)–(9) become, in terms of the corresponding Hallén quantities $M_n(z)_H$,

$$M_n(z)_G = \sum_{i=0}^n \frac{n!}{(n-i)!} (\bar{\Omega}_G - \Omega)^i M_{n-i}(z)_H, \quad n \geq 0, \quad (33)$$

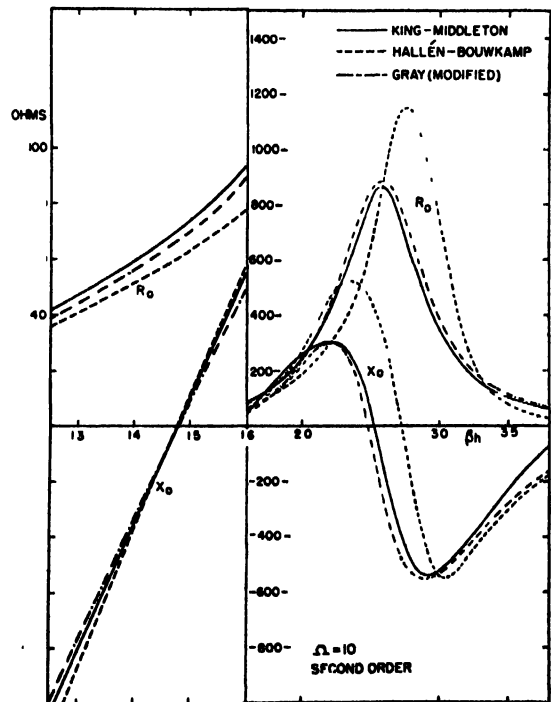


FIG. 3. Second-order impedances; $\Omega = 10$.

where it is understood that

$$M_0(z)_G = M_0(z)_H \equiv \sin \beta(h - |z|). \quad (34)$$

The m th order current is obtained in the form (3) after a suitable rearrangement of terms. It becomes finally

$$(I_n)_{mG} = \frac{j2\pi V_0^*}{R_c \bar{\Omega}_G} \left\{ \frac{(D_{1G})_m \sin \beta(h - |z|) + \sum_{n=1}^m (D_{n+1G})_m M_n(z)_H / \bar{\Omega}_G^n}{\cos \beta h + \sum_{n=1}^m (D_{nG})_{m-1} F_n(h)_H / \bar{\Omega}_G^n} \right\}, \quad (35)$$

where with $y \equiv 1 - \Omega / \bar{\Omega}_G$, a table similar to (75), reference 10, may be constructed (see Table I). For a finite number of terms, $m < \infty$, all terms in Table I belonging to a given order m of solution must be retained, and no others. Thus, if an m th-order solution is evaluated, only terms contributed by $M_n(z)_G$, $F_n(h)_G$, where $n = 0, 1, \dots, m$ are used. Hence for a first-order theory ($m = 1$)

$$(D_{1G})_0 = 1, \quad (D_{1G})_1 = 1 + \left(1 - \frac{\Omega}{\bar{\Omega}_G}\right), \quad (36)$$

$$(D_{2G})_1 = 1$$

only are needed in (35), and for a second-order expansion ($m = 2$) only

$$(D_{1G})_2 = 1 + (1 - \Omega / \bar{\Omega}_G) + (1 - \Omega / \bar{\Omega}_G)^2; \quad (37)$$

$$(D_{2G})_2 = 1 + 2(1 - \Omega / \bar{\Omega}_G); \quad (D_{3G})_2 = 1$$

are required. Curves of the quantities in (36) and (37) are shown in Fig. 6. When an infinite number of terms ($m \rightarrow \infty$) are summed, one observes that

$$(D_{nG})_{m \rightarrow \infty} = \frac{1}{(n-1)! dy^{n-1}} \left(\frac{1}{1-y} \right) = \frac{1}{(1-y)^n} = (\bar{\Omega}_G / \Omega)^n, \quad y < 1, n \geq 1. \quad (38)$$

Then, with the factors $(D_{nG})_\infty$, (35) becomes identical with the Hallén-Bouwkamp solution for the current and impedance. It is further evident that if the radius a of the antenna becomes vanishingly small, $\bar{\Omega}_G$ approaches Ω for all values of βh , so that (35) also approaches the Hallén-Bouwkamp result. The impedance is readily obtained from (12) and (35). It is plotted in Figs. 1-4 for first and second orders. A comparison of the above with Sections V and VI of reference (10) shows that the modified Gray and King-Middleton solutions are formally identical;

however, $\bar{\Omega}_G$ and $\Psi_{\text{King-Mid}}$ are in general numerically distinct because of the different assumptions anent the distribution function $g(z, z')$.

(c) Unmodified Gray Solution⁸

The expansion for the current in this case is obtained with the same value of $g(z, z')$ as for the modified development, cf. Eq. (22), and yields the same expansion parameter $\bar{\Omega}_G$, Eq. (26). However, the generalized formulas for $(F_n)_G$ and $(G_n)_G$ are

$$F_n(z)_G = F_{n-1}(z)_G \int_{-h}^h g(z, z')_G \frac{\exp(-j\beta R_1)}{R_1} dz' - \int_{-h}^h F_{n-1}(z')_G \frac{\exp(-j\beta R_1)}{R_1} dz', \quad (39)$$

with similar expressions for $G_n(z)_G$. Here the subscript g denotes the unmodified Gray solution. Note that terms involving $F_n(h)_G$ and $G_n(h)_G$ are missing. This occurs because the expression for the current at $z = \pm h$ has not been subtracted from both sides of the fundamental integral Eq. (1) to insure the vanishing of the current at the ends of the antenna for all values of βh . For example, when $\beta h = \pi/2$, the arbitrary constant C_1 does not appear in (1) in its present form when $z = \pm h$, and it is not then possible to determine

TABLE I.

Order $m =$	0	1	2	3	4	...
$(D_{1G})_m$	1	$+y$	$+y^2$	$+y^3$	$+y^4$	$+\dots$
$(D_{2G})_m$		1	$+2y$	$+3y^2$	$+4y^3$	$+\dots$
$(D_{3G})_m$			1	$+3y$	$+6y^2$	$+\dots$
$(D_{4G})_m$				1	$+4y$	$+\dots$
$\vdots$						
$(D_{nG})_m$						

C_1 to satisfy the boundary conditions at the ends. As before, it is convenient for purposes of comparison and computation to express the unmodified Gray functions in terms of the (ex-

tensively) tabulated Hallén quantities. The results are given in the Appendix.

The m th order current distribution becomes, in the present instance

$$(I_z)_{m0} = \frac{j2\pi V_0}{R_0 \Psi_0(z)} \left\{ \frac{\sin \beta(h - |z|) + \sum_{n=1}^m M_n(z)_0 / \bar{\Omega}_0^n}{\cos \beta h + \sum_{n=1}^m F_n(h)_0 / \bar{\Omega}_0^n} \right\}, \quad (40)$$

where from (5) and (22)

$$\Psi(z)_0 = \int_{-h}^h \frac{\cos \beta R_1}{R_1} dz' = \Psi(z)_H - \text{Cin } \beta(h+z) - \text{Cin } \beta(h-z). \quad (41)$$

Although taking $\langle \Psi(z)_0 \rangle_H = \bar{\Omega}_0$ as the expansion parameter, Gray *does not replace* $\Psi(z)_0$ in (40) by its average value. From (12) the m th order impedance is seen to be

$$(Z_0)_{m0} = -j60(\Omega - 2 \text{Cin } \beta h) \left\{ \frac{\cos \beta h + \sum_{n=1}^m F_n(h)_0 / \bar{\Omega}_0^n}{\sin \beta h + \sum_{n=1}^m M_n(0)_0 / \bar{\Omega}_0^n} \right\}, \quad (42)$$

since $R_0/2\pi = 60$ ohms. For a first-order solution (42) gives

$$(Z_0)_{10} = -j60(\Omega - 2 \text{Cin } \beta h) \left\{ \frac{\cos \beta h + F_1(h)_0 / \bar{\Omega}_0}{\sin \beta h + M_1(0)_0 / \bar{\Omega}_0} \right\}, \quad (43)$$

while Gray chooses for her solution a combination of first- and second-order terms, to wit,

$$(Z_0)_{30/2} = -j60(\Omega - 2 \text{Cin } \beta h) \left\{ \frac{\cos \beta h + F_1(h)_0 / \bar{\Omega}_0}{\sin \beta h + M_1(0)_0 / \bar{\Omega}_0 + [G_1(h)_0 F_1(0)_0 - F_1(h)_0 G_1(0)_0] / \bar{\Omega}_0^2} \right\}. \quad (44)$$

This we call the three-halves order solution, inasmuch as the terms in $F_2(h)$, $G_2(h)$, $F_2(0)$, $G_2(0)$ are missing. Curves illustrating (43) and (44) for $\Omega = 10$, $\Omega = 20$ are shown in Figs. 7 and 8.

(d) King-Middleton Solution¹⁰

Here the relative distribution function $g(z, z')_K$ is

$$g(z, z')_K = I_{z'}/I_z = \sin \beta(h - |z'|) / \sin \beta(h - |z|), \quad (45)$$

representing the ratio of the *approximate current distributions*, which to the zeroth order are sinusoidal, and to the first order are nearly so.^{8,10} The expansion parameter becomes

$$\Psi = \psi = |\Psi(0)_K| = |\psi(0)| / \sin \beta h, \quad \beta h \leq \pi/2 \quad (46)$$

$$= |\Psi(h - \lambda/4)_K| = |\psi(h - \lambda/4)|, \quad \beta h \geq \pi/2.$$

The quantity ψ is plotted in Fig. 5 together with $\bar{\Omega}_0$. The point $z = z_0$ of Eq. (4) is thus seen to be zero when $\beta h \leq \pi/2$ and $h - \lambda/4$ for $\beta h \geq \pi/2$. The

imaginary part of $\Psi(z)_K$ is always sufficiently small compared with the real part that it may be neglected, except near $z = \pm h$, with the result that $\gamma(z)_K \doteq 0$. A detailed account of the derivation of ψ is given in Section IV, reference 10.

A useful formula for small arguments is

$$\Psi(0)_K = \Omega - 2 - j\beta h, \quad (47a)$$

$$\psi = |\Psi(0)_K| = \Omega - 2 + \frac{1}{2}\beta h / (\Omega - 2) \quad (47b)$$

The function $60(\Omega - 2)$ is the characteristic impedance used by Schelkunoff.⁹ It is seen to be a good approximation for *short* antennas.

As mentioned in (b) the results of the King-Middleton solution are formally identical with those of the modified Gray solution, Eqs. (31)–(38), provided $\gamma = 1 - \Omega/\bar{\Omega}_0$ is replaced by $x = 1 - \Omega/\psi$, and $(D_n)_m$ is altered accordingly. The actual results are noticeably different, as can be seen from Figs. 1–4, since ψ and $\bar{\Omega}_0$ are derived using different assumptions.

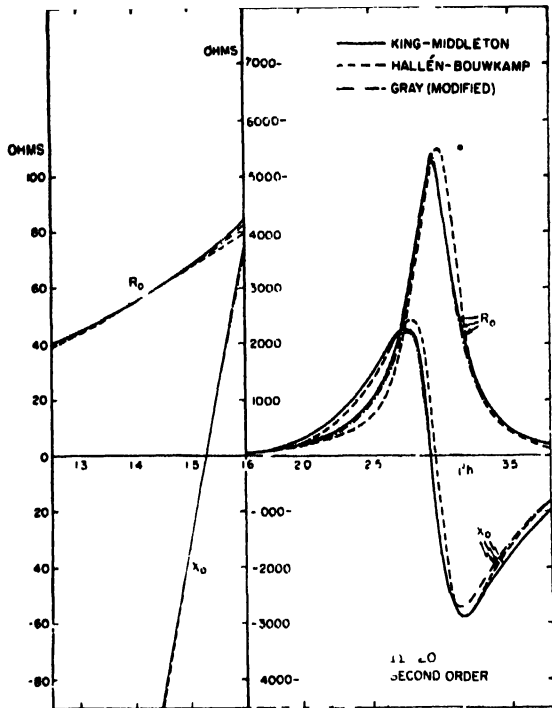


FIG. 4. Second-order impedances; $\Omega = 20$.

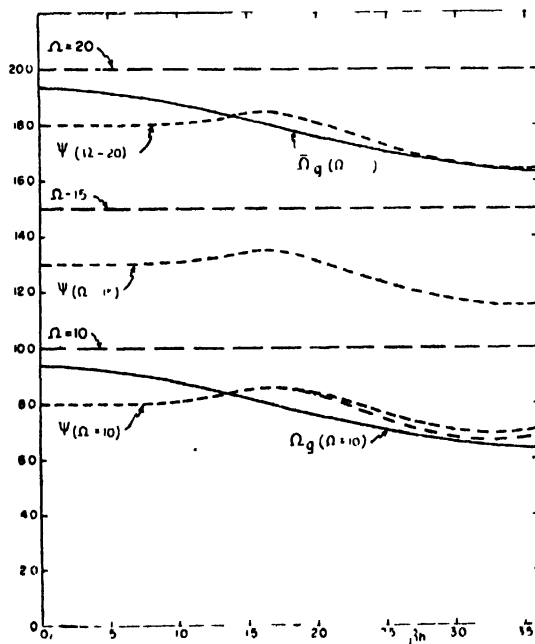


FIG. 5. $\bar{\Omega}_g$ and ψ as functions of βh .

III. COMPARISON OF THEORIES AND EXPERIMENT

The broad analytical comparison of the theories of Hallén and Bouwkamp, of Gray,

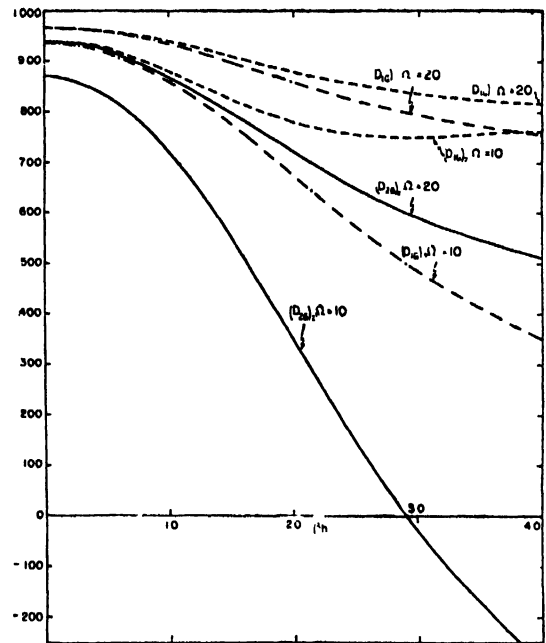


FIG. 6. The functions D_{1g} and D_{2g} defined in Table I.

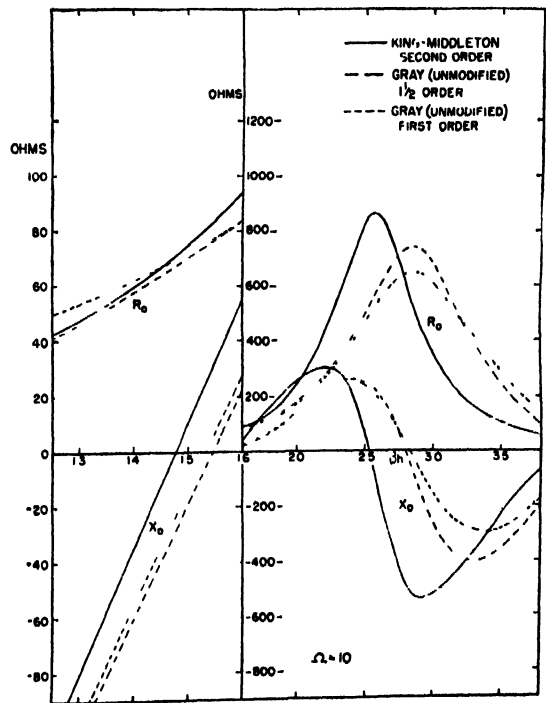


FIG. 7. Impedances according to Gray, $\Omega = 10$.

unmodified and modified, and of King and Middleton, reveals that they all solve the identical *mathematical* problem. In all cases the antenna analyzed is the "theoretical antenna"

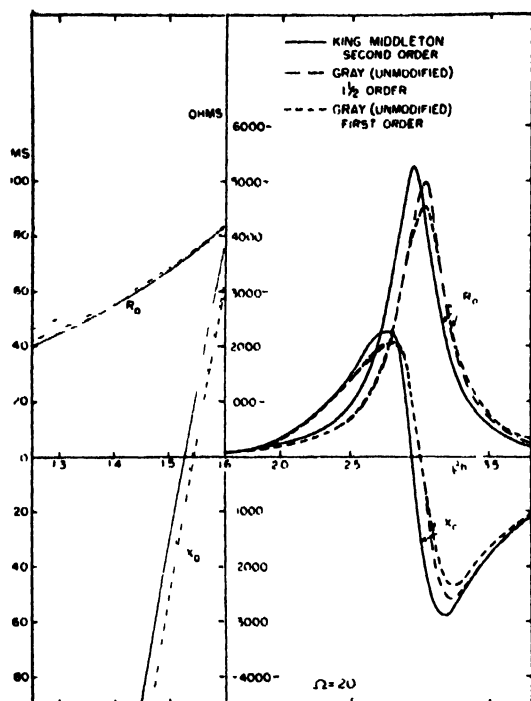


FIG. 8. Impedances according to Gray, $\Omega = 20$.

with its physically unavailable "slice generator," as discussed in an earlier paper.² All of the analyses are one-dimensional, all concern themselves with the same integral equation originally set up by Hallén, all follow essentially the same method of expansion in a series. They differ fundamentally only in the choice of function selected to describe approximately the distribution of current under the sign of integration and in the resulting parameter of expansion. All theories give the same results for infinitely thin antennas ($\Omega = \infty$), and all except the unmodified Gray reduce to the same form if an infinite number of terms is used. The unmodified Gray solution differs in this one respect as a result of a peculiar inconsistency in the use of the expansion parameter.

From the practical point of view that theory is most satisfactory which converges most rapidly, other things being equal. In particular it should converge sufficiently rapidly so that at least a second-order solution is a good approximation. Because of the intricacy of the series it has thus far not been possible to determine the rapidity of convergence by direct tests. There remains, however, the obvious fact that the

better the original approximation for the current under the sign of integration the more rapid must be the convergence. On this basis the King-Middleton theory is unique in that *it alone actually approximates the distribution of current* in expanding the integral. In effect, the Hallén-Bouwkamp analysis assumes a current of uniform amplitude along the entire antenna equal to that at the point of calculation. It also neglects retardation. The modified and unmodified Gray theories assume a distribution given by $\cos \beta R$ with R measured from the point of calculation. Retardation is also neglected. Neither of these assumptions is even remotely related to the *actual* distribution. The King-Middleton theory, on the other hand, uses in the integrand the actual zeroth order distribution of current common to all theories and also takes retardation into account. The Hallén-Bouwkamp theory is the simplest and the crudest because the expansion parameter Ω turns out to be independent of βh . The Gray formulation, while it represents the current no better than does the Hallén form, leads to an expansion parameter that is a function of βh as well as of h/a and thereby achieves a better approximation. Finally, the King-Middleton theory, by fitting the current quite accurately, obtains a still better expansion parameter that is actually an excellent approximation of a quantity proportional to the magnitude of the ratio of vector potential to current along the antenna. This ratio has only a very small imaginary part and is sensibly constant over the entire length of the antenna except near and at the ends, where it becomes infinite. Of the three solutions, the King-Middleton form is without question the most refined and it should, therefore, be in better agreement with the results of properly arranged experiments.

Before considering a comparison of theory with experiment, a few remarks must be made regarding the *unmodified* Gray solution. This was presumably originally intended to bring Hallén's theory into better agreement with the theory of Schelkunoff. In view of this purpose, Miss Gray did not take full advantage of the superiority of her parameter of expansion over that of Hallén. In particular, while introducing an *average value* of the expansion parameter in all integrals, she retained the value *at the origin*

as a principal common factor in the expression for the impedance. Thus the unmodified Gray solution actually uses *two* parameters, whereas the modified Gray form uses consistently the average value throughout. As a further but much less important difference, in the modified Gray form, the solution at $z=h$ is subtracted from that at $z=0$. This was not done by Miss Gray. Superficially, this subtraction appears to be an important and necessary step since without it the constant of integration disappears from the solution whenever $\beta h = \frac{1}{2}(2n+1)\pi$, $n=0, 1, 2, \dots$. Actually, in the expanded form, the constant of integration occurs in higher order terms, so that its evaluation is nevertheless possible. A final and seemingly quite unjustified step in the original Gray solution is the addition of only a part of the second-order term to the first-order solution *without any investigation of the relative magnitudes of the two parts*. The term so retained (in the words of Miss Gray) "is not negligible numerically compared with the preceding [first-order] term and it seems advisable to make the current as nearly zero as possible at the ends of the antenna." Actually there is no reason whatever to suppose that the omitted part of the second-order term is negligible or even smaller than the part retained. Indeed, it might possibly be the negative of the part retained. The argument that part of the term must be kept "to make the current as nearly zero as possible at the ends of the antenna" is peculiarly inapplicable because the current as given by Miss Gray's own formula (Eq. (11), reference 8) *is always zero* at the ends with or without the term.

In order to facilitate intercomparison of the several theories, curves showing anti-resonant and resonant resistances, anti-resonant and resonant shifts in βh , respectively, from π and $\pi/2$, and resistances and reactances at $h=\lambda/4$, are given in Figs. 9–13 for both first- and second-order solutions. In the unmodified Gray theory no complete second-order solution is available. Instead the solution of Miss Gray involving only half of the second-order solution is given and marked $1\frac{1}{2}$ order. Its significance is questionable until the complete second order is determined.

First- and second-order solutions differ from each other not only in magnitude of resistance and reactance but also in the magnitude of the

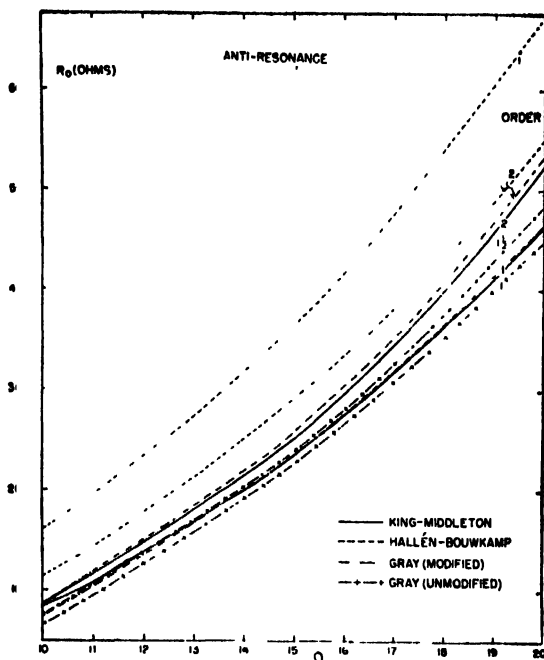


FIG. 9. Anti-resonant resistances as a function of Ω .

shift to shorter lengths of the entire impedance curves. As a result, a simple and clear-cut comparison of convergence by determining the difference between first and second order values is meaningful only for the magnitudes of the resistances at anti-resonance. From Fig. 9 it is clear that first- and second-order solutions agree quite closely in the King-Middleton theory, almost as well in the *modified* Gray theory, and rather poorly in the Hallén-Bouwkamp solution. This is as predicted above from a consideration of the refinements of the several theories. At $\Omega=15$, for example, the difference between first- and second-order anti-resonant resistances is approximately 6 percent for the King-Middleton theory, 9 percent for the modified Gray theory, and 24 percent for the Hallén-Bouwkamp theory when referred to the second-order values. The difference between first and $1\frac{1}{2}$ -order solutions in the unmodified Gray theory of course reveals nothing about convergence.

As pointed out in discussing the correlation of experiment with theory,² there is no physically available equivalent of the idealized slice generator actually assumed or implied in all of the theoretical analyses outlined above. If a continuous conductor is loosely and symmetrically coupled to a loop at the end of a two-wire line,

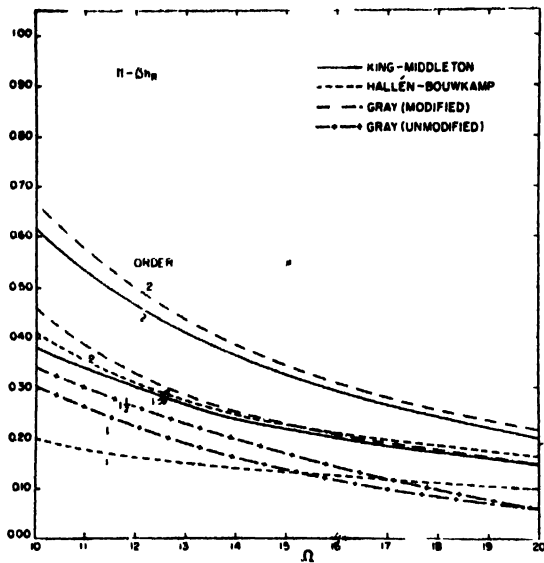


FIG. 10. Difference in radians between the anti-resonant length π of an infinitely thin antenna ($\Omega = \infty$) and the anti-resonant lengths of antennas in the range of Ω from 10 to 20.

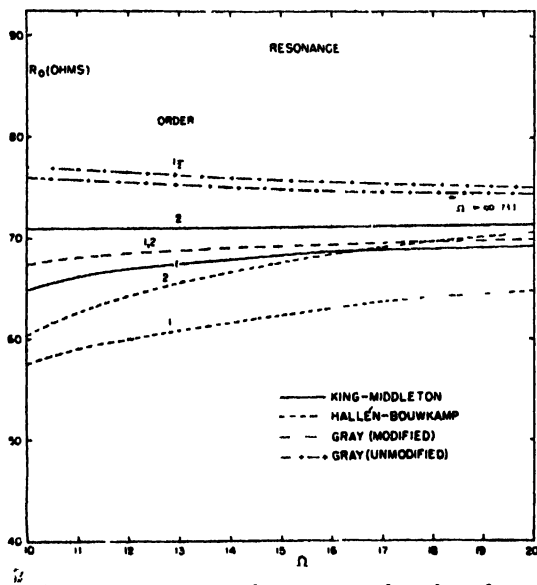


FIG. 11. Resonant resistances as a function of Ω .

the e.m.f. induced in the wire is exactly equivalent to a "slice" generator at the center of the conductor. Similarly, the lumped e.m.f. in the equivalent series circuit of a receiving antenna is in effect that of a "slice" generator. In the case of direct drive probably the nearest physical equivalent of such a generator is a base-driven, cylindrical antenna with hemispherical ends erected vertically over a highly conducting metal plane of sufficient size, the antenna being

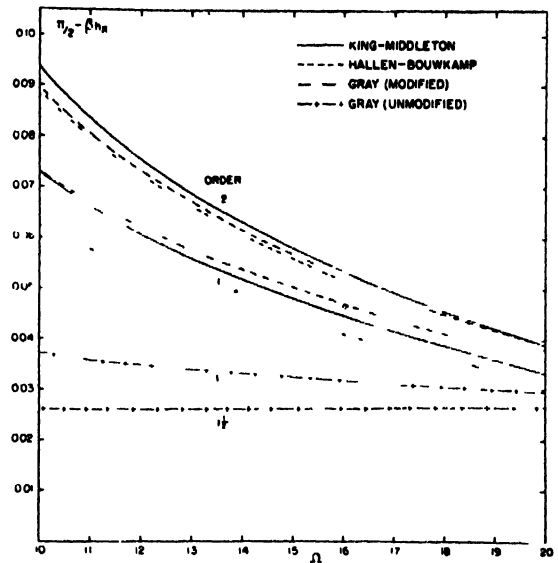


FIG. 12. Difference in radians between the resonant length $\pi/2$ of an infinitely thin antenna ($\Omega = \infty$) and the resonant lengths of antennas in the range of Ω from 10 to 20.

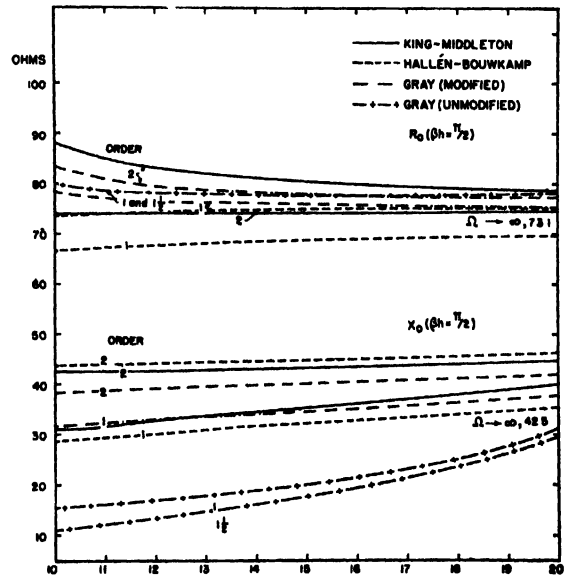


FIG. 13. Resistances and reactances of the half-wave dipole, $\beta h = \pi/2$, as a function of Ω .

the extension of the inner conductor of a coaxial line of very small cross section.

Very careful measurements have been made by D. D. King on a coaxial arrangement at a frequency of about 750 megahertz. These are described in complete detail in a doctoral dissertation and will be submitted for publication shortly. A cross section of the experimental results for an antenna with $h/a = 75$ or $\Omega = 10$ is

given in Table II in the form of a list of the primary critical values. The corresponding values for the several theories discussed above are also shown, together with the percentage departure from the experimental values.

In addition, corresponding critical values are given for Schelkunoff's transmission-line theory.⁹ The data in this case were estimated from Figs. 23-26 of reference 9. Unfortunately the ratio $|X|_{\min}/|X|_{\max}$ was not actually available in this case since $|X|_{\min}$ does not appear on the curves and extrapolation is difficult. However, the relation $|X|_{\min} + |X|_{\max} \doteq R_{\max}$ is generally a good approximation in all of the theories, so that this relation was used to determine an approximate value of $|X|_{\min}$. Evidently it may be considerably in error, and no great weight should be attached to it until confirmed.

Examination of Table II shows clearly that by and large the King-Middleton second-order theory agrees best with the experimental values, the modified Gray results are next best, the Hallén-Bouwkamp values are poorest. The unmodified Gray and the Schelkunoff values agree quite satisfactorily with this experimental data for resistance at anti-resonance, but show very wide departures for other critical values. On the other hand, the King-Middleton second-order theory agrees very much better with the experimental data for anti-resonant resistance supplied by Schelkunoff than does his own theory. This is shown in Fig. 14 where Schelkunoff's theoretical curve (broken line), his experimental points (small circles), and the King-Middleton curve (unbroken line) are shown together as a function of Ω . In view of the great dependence of the impedance on the driving conditions, as shown for example in the work of Brown and

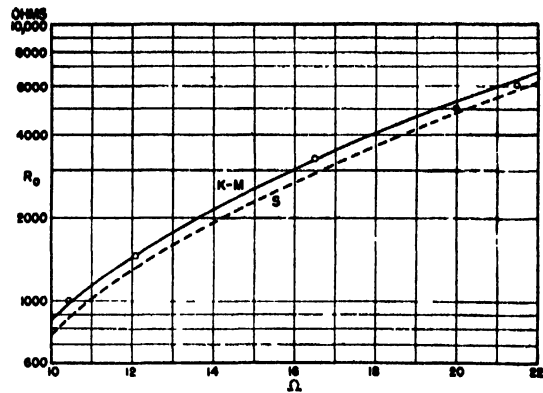


FIG. 14. Resistance at anti-resonance as a function of Ω . The solid-line curve marked K-M is according to the King-Middleton second-order theory; the broken-line curve is according to the transmission-line theory of S. A. Schelkunoff. The broken-line curve and the experimental points indicated by small circles are taken from S. A. Schelkunoff, Proc. I.R.E. 29, 511, Fig. 26 (September, 1941).

Woodward,¹² and the impossibility of experimentally reproducing as a practical driving device the theoretical slice generator, care must be exercised not to overemphasize the significance of the close agreement. It is, nevertheless, safe and gratifying to conclude that the analytically elegant integral-equation method of Hallén leads to a highly refined and satisfactory theory of the thin cylindrical antenna.

APPENDIX

With the help of the appendix to reference (10), it is seen that

$$\int_{-h}^h \frac{\cos \beta R_1}{R_1} dz' \doteq \Psi(z)_H - \text{Cin } \beta(h+z) - \text{Cin } \beta(h-z), \quad h^2 \gg a^2, \quad (i)$$

and

$$\int_{-h}^h \frac{\exp(-j\beta R_1)}{R_1} dz' \doteq \Psi_H(z) - \text{Cin } \beta(h+z) - \text{Cin } \beta(h-z) - j \text{Si } \beta(h+z) - j \text{Si } \beta(h-z), \quad h^2 \gg a^2. \quad (ii)$$

From (i) and (ii) the following results, useful in the com-

TABLE II. Critical values related to the impedance characteristic of a cylindrical antenna for which $\Omega = 2 \ln(2h/a) = 10$

	Experimental data from Ph.D. Thesis of D. D. King	King-Middleton second-order theory	Modified Gray second-order theory	Hallén-Bouwkamp second-order theory	Unmodified Gray 1½ order theory	Schelkunoff transmission- line theory
$(R_0)_{\text{anti res.}}$	800	860	885	1150	740	740
$\pi - \beta h_{\text{anti res.}}$	0.60	0.614	0.665	0.41	0.34	0.48
$X_{\min}/X_{\max}$	1.95	1.8	1.84	1.06	1.54	(1.4) (28)
$(R_0)_{\text{res.}}$	71.5	71.0	67	60.6	77	61.4
$(\pi/2) - \beta h_{\text{res.}}$	0.098	0.094	0.0895	0.0885	0.026	0.094
$(R_0)_{\beta h = \pi/2}$	85	88	83.5	73.5	80	75
$(X_0)_{\beta h = \pi/2}$	47	42.5	38	43.5	12	45
		%	%	%	%	%
		7.5	13.5	47.5	7.5	7.5
		2.3	10.9	46.3	43.4	20
		7.7	5.7	46	21.0	14.1
		0.7	6.3	15.2	7.7	6.1
		6.1	8.2	10.3	73.5	18
		3.5	1.8	13.6	6.3	4.2
		9.5	19.1	5.3	74.5	

putation of the current distribution, may be obtained:

$$F_1(z)_0 = F_1(z)_H + (\cos \beta z - \cos \beta h) [\Psi(z)_H - \Omega - \text{Cin } \beta(h+z) - \text{Cin } \beta(h-z)] + j \cos \beta h [Si \beta(h+z) + Si \beta(h-z)], \quad (\text{iii})$$

and

$$G_1(z)_0 = G_1(z)_H + (\sin \beta |z| - \sin \beta h) [\Psi(z)_H - \Omega - \text{Cin } \beta(h+z) - \text{Cin } \beta(h-z)] + j \sin \beta h [Si \beta(h+z) + Si \beta(h-z)]. \quad (\text{iv})$$

Important special cases are

$$F_1(0)_0 = F_1(0)_H + 2(\cos \beta h - 1) \text{Cin } \beta h + 2j \cos \beta h Si \beta h, \quad (\text{v})$$

$$F_1(h)_0 = F_1(h)_H + j \cos \beta h Si 2\beta h, \quad (\text{vi})$$

and

$$G_1(0)_0 = -Si 2\beta h + j \text{Cin } 2\beta h, \quad (\text{vii})$$

$$G_1(h)_0 = \sin \beta h [\Psi(h)_H - \text{Cin } 2\beta h] - Si(h), \quad (\text{viii})$$

where

$$\Psi(h)_H = \sinh^{-1} \frac{2h}{a}, \quad \Psi(0)_H = 2 \sinh^{-1} \frac{h}{a}, \quad (\text{ix})$$

and

$$S_1(h) = \frac{\sin \beta h}{2} \{ \text{Cin } 4\beta h - 2 \text{Cin } 2\beta h \} + \frac{\cos \beta h}{2} \{ Si 4\beta h - 2 Si 2\beta h \} - j \frac{\cos \beta h}{2} \{ \text{Cin } 4\beta h - 2 \text{Cin } 2\beta h \} - j \sin \beta h \cdot Si 2\beta h + j \frac{\sin \beta h}{2} Si 4\beta h. \quad (\text{x})$$

With the help of the above one may write

$$M_1(0)_0 = M_1(0)_H + \sin \beta h \{ j Si 2\beta h - 2 \text{Cin } \beta h \}. \quad (\text{xi})$$

$M_1(0)_H$, $F_1(0)_H$, $F_1(h)$ are tabulated in reference 4.

Calendar of Meetings

April		10-11	Acoustical Society of America, New York, New York
1- 4	American Society of Mechanical Engineers, Chattanooga, Tennessee	27-29	American Geophysical Union, Washington, D. C.
3- 5	Society of Automotive Engineers, New York, New York	June	
8-12	American Chemical Society, Atlantic City, New Jersey	2- 5	American Society of Refrigerating Engineers, St. Paul, Minnesota
12-13	American Physical Society, Southeastern Section, Atlanta, Georgia	13 14	Institute of Aeronautical Sciences, Detroit, Michigan
17-19	American Society of Civil Engineers, Philadelphia, Pennsylvania	17-20	American Society of Mechanical Engineers, Detroit, Michigan
18-20	American Philosophical Society, Philadelphia, Pennsylvania	20-23	Society for the Promotion of Engineering Education, St. Louis, Missouri
22-24	National Academy of Sciences, Washington, D. C.	24 28	American Institute of Electrical Engineers, Detroit, Michigan
25-27	American Physical Society, Cambridge, Massachusetts	24-28	American Society for Testing Materials, Buffalo, New York
26-27	American Mathematical Society, Chicago, Illinois		
26-27	American Mathematical Society, New York, New York		
27	American Mathematical Society, West Coast	July	
28 May 1	American Ceramic Society, Inc., Buffalo, New York	17-19	American Society of Civil Engineers, Spokane, Washington
May		August	
6-10	Society of Motion Picture Engineers, New York, New York	14-15	Institute of Aeronautical Sciences, Los Angeles, California
		19-20	Mathematical Association of America, Ithaca, New York

Correction of Diffraction Amplitudes for Lorentz and Polarization Factors

M. J. BUEGER AND GILBERT E. KLEIN

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received December 13, 1945)

The corrections for Lorentz and polarization factors for trial and error crystal structure computations are discussed, following the general method recently presented for correcting x-ray diffraction data used in Fourier syntheses. The Lorentz correction has the form $(1/S) \csc T$, where S is a scale constant characteristic of the reciprocal lattice level and whose form varies with the method of recording the data. The scale constant is eliminated by appropriate timing of the exposure. When many trial and error computations have to be made, there are advantages in using a scheme involving a correction factor $(\csc T)^{\frac{1}{2}}$ instead of $\csc T$. Tables are provided for these functions, for the polarization correction and for the combined Lorentz-polarization correction which is used in the special case of zero-level photographs. Tables are also provided for the functions $(h^2 + k^2)^{\frac{1}{2}}$ and $(h^2 + k^2 + hk)^{\frac{1}{2}}$, which are useful for computing the argument for the zero-level corrections of tetragonal and hexagonal crystals.

INTRODUCTION

IN an earlier contribution,¹ the correction of x-ray diffraction intensities for Lorentz and polarization factors was discussed, and appropriate tables were presented for carrying out the correction. This procedure is important in deriving the F 's required for any of the several Fourier series methods of investigating crystal structures. In this paper we discuss the inverse problem, namely, the computation of expected x-ray diffraction intensities from a set of diffracted amplitudes, F . This computation is continually dealt with in fixing the parameters of atoms by the method of trial and error.

The theory of Lorentz and polarization corrections has already been discussed in detail.¹ We merely outline its essential features here. The relation between F and the required intensity I is

$$I \sim TLpF^2, \quad (1)$$

where T is the time of the exposure, L is the Lorentz factor, and p is the polarization factor.

The polarization factor, p , is tabulated in Table I for the argument $\sin \theta$, and in Table II for the argument σ (the distance of the reciprocal lattice point from the origin). For notes on the computation of σ , the reader is referred to the earlier paper.¹

THE LORENTZ FACTOR

For diffraction techniques which make use of a single crystal and which separate the diffraction

record into layer lines, the Lorentz factor has the form

$$L = \frac{1}{S} \cdot \frac{1}{\sin T}, \quad (2)$$

where T is an angular coordinate of the reflection,² and S is a scale factor which is a function of the reciprocal lattice level coordinate and which varies in form¹ according to the recording method. For the equi-inclination Weissenberg method,

$$S = \cos^2 \mu = (1 - [\xi/2]^2)^{\frac{1}{2}}. \quad (3)$$

If (1) and (2) are combined, and if the total time of exposure, T , is split into T_0 (the time of exposure chosen for the zero level) and t (a numerical factor), there results

$$I \sim T_0 \left[\frac{1}{t} \right] \frac{1}{S \sin T} p |F|^2. \quad (4)$$

It is evident that the time of exposure, T_0 , can be selected for each level so that the product in brackets becomes unity. This merely requires that

$$t = S. \quad (5)$$

For the equi-inclination Weissenberg method, according to (3), the timing factor becomes

$$t = \cos^2 \mu. \quad (6)$$

This procedure amounts to elimination of the scale part of the Lorentz factor by time cancella-

¹ M. J. Buerger and Gilbert E. Klein, "Correction of x-ray diffraction intensities for Lorentz and polarization factors," *J. App. Phys.* 16, 408-418 (1945).

² M. J. Buerger, *X-Ray Crystallography* (John Wiley & Sons, Inc., New York, 1942), p. 222 *et seq.*

TABLE I. Polarization factor, $p \left(= \frac{1 + \cos^2 2\theta}{2} \right)$, as a function of $\sin \theta$.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.00	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9998
.01	.9998	.9998	.9997	.9997	.9996	.9996	.9995	.9994	.9994	.9993
.02	.9992	.9991	.9990	.9989	.9988	.9988	.9987	.9985	.9984	.9983
.03	.9982	.9981	.9980	.9978	.9977	.9976	.9974	.9973	.9971	.9970
.04	.9968	.9966	.9965	.9963	.9961	.9960	.9958	.9956	.9954	.9952
.05	.9950	.9948	.9946	.9944	.9942	.9940	.9937	.9935	.9933	.9931
.06	.9928	.9926	.9923	.9921	.9918	.9916	.9913	.9911	.9908	.9905
.07	.9902	.9900	.9897	.9894	.9891	.9888	.9885	.9882	.9879	.9876
.08	.9873	.9870	.9866	.9863	.9860	.9857	.9853	.9850	.9846	.9843
.09	.9839	.9836	.9832	.9828	.9825	.9821	.9817	.9814	.9810	.9806
.10	.9802	.9798	.9794	.9790	.9786	.9782	.9778	.9774	.9769	.9765
.11	.9761	.9757	.9752	.9748	.9743	.9739	.9735	.9730	.9725	.9721
.12	.9716	.9711	.9707	.9702	.9697	.9692	.9688	.9683	.9678	.9673
.13	.9668	.9663	.9658	.9652	.9647	.9642	.9637	.9632	.9626	.9621
.14	.9616	.9610	.9605	.9599	.9594	.9588	.9583	.9577	.9571	.9566
.15	.9560	.9554	.9549	.9543	.9537	.9531	.9525	.9519	.9513	.9507
.16	.9501	.9495	.9489	.9483	.9477	.9470	.9464	.9458	.9451	.9445
.17	.9439	.9432	.9426	.9419	.9413	.9406	.9400	.9393	.9386	.9380
.18	.9373	.9366	.9360	.9353	.9346	.9339	.9332	.9325	.9318	.9311
.19	.9304	.9297	.9290	.9283	.9276	.9269	.9261	.9254	.9247	.9239
.20	.9232	.9225	.9217	.9210	.9202	.9195	.9187	.9180	.9172	.9164
.21	.9157	.9149	.9142	.9134	.9126	.9118	.9110	.9102	.9095	.9087
.22	.9079	.9071	.9063	.9055	.9047	.9039	.9031	.9023	.9014	.9006
.23	.8998	.8990	.8981	.8973	.8965	.8956	.8948	.8940	.8931	.8923
.24	.8914	.8906	.8897	.8888	.8880	.8872	.8863	.8854	.8846	.8837
.25	.8828	.8819	.8811	.8802	.8793	.8784	.8775	.8766	.8757	.8748
.26	.8739	.8730	.8721	.8712	.8703	.8694	.8685	.8676	.8667	.8658
.27	.8648	.8639	.8630	.8621	.8611	.8602	.8593	.8583	.8574	.8564
.28	.8555	.8545	.8536	.8526	.8517	.8507	.8498	.8488	.8479	.8469
.29	.8459	.8450	.8440	.8430	.8420	.8411	.8401	.8391	.8382	.8372
.30	.8362	.8352	.8342	.8332	.8323	.8313	.8303	.8293	.8283	.8273
.31	.8263	.8253	.8243	.8233	.8223	.8212	.8202	.8192	.8182	.8172
.32	.8162	.8152	.8141	.8131	.8121	.8111	.8100	.8090	.8080	.8070
.33	.8059	.8049	.8039	.8028	.8018	.8007	.7997	.7987	.7976	.7966
.34	.7955	.7944	.7934	.7924	.7913	.7903	.7892	.7882	.7871	.7861
.35	.7850	.7840	.7829	.7818	.7808	.7797	.7787	.7776	.7765	.7754
.36	.7744	.7733	.7723	.7712	.7701	.7690	.7680	.7669	.7658	.7648
.37	.7637	.7626	.7615	.7605	.7594	.7583	.7572	.7561	.7551	.7540
.38	.7529	.7518	.7507	.7497	.7486	.7475	.7464	.7453	.7442	.7432
.39	.7421	.7410	.7399	.7388	.7377	.7366	.7356	.7345	.7334	.7323
.40	.7312	.7301	.7290	.7279	.7268	.7258	.7247	.7236	.7225	.7214
.41	.7203	.7192	.7182	.7170	.7160	.7149	.7138	.7127	.7116	.7105
.42	.7094	.7083	.7073	.7062	.7051	.7040	.7029	.7018	.7007	.6997
.43	.6986	.6975	.6964	.6953	.6942	.6932	.6921	.6910	.6899	.6888
.44	.6878	.6867	.6856	.6845	.6835	.6824	.6813	.6802	.6792	.6781
.45	.6770	.6759	.6749	.6738	.6727	.6717	.6706	.6695	.6685	.6674
.46	.6663	.6653	.6642	.6632	.6621	.6611	.6600	.6589	.6579	.6568
.47	.6558	.6547	.6537	.6527	.6516	.6506	.6495	.6485	.6474	.6464
.48	.6454	.6443	.6433	.6423	.6412	.6402	.6392	.6382	.6371	.6361
.49	.6351	.6341	.6331	.6320	.6310	.6300	.6290	.6280	.6270	.6260
.50	.6250	.6240	.6230	.6220	.6210	.6200	.6190	.6180	.6170	.6161
.51	.6151	.6141	.6132	.6122	.6112	.6102	.6093	.6083	.6073	.6064
.52	.6054	.6044	.6035	.6026	.6016	.6006	.5997	.5988	.5979	.5969
.53	.5960	.5951	.5942	.5932	.5923	.5914	.5905	.5896	.5887	.5878
.54	.5869	.5860	.5851	.5842	.5833	.5824	.5815	.5806	.5798	.5789
.55	.5780	.5771	.5763	.5754	.5746	.5737	.5729	.5720	.5712	.5703
.56	.5695	.5687	.5678	.5670	.5662	.5654	.5645	.5637	.5629	.5621
.57	.5613	.5605	.5597	.5589	.5582	.5574	.5566	.5558	.5551	.5543
.58	.5535	.5528	.5520	.5513	.5505	.5498	.5490	.5483	.5476	.5469
.59	.5461	.5454	.5447	.5440	.5433	.5426	.5419	.5412	.5406	.5399
.60	.5392	.5385	.5379	.5372	.5365	.5359	.5353	.5346	.5340	.5333
.61	.5327	.5321	.5315	.5309	.5303	.5297	.5291	.5285	.5279	.5273
.62	.5267	.5262	.5256	.5250	.5245	.5239	.5234	.5228	.5223	.5218
.63	.5213	.5207	.5202	.5197	.5192	.5187	.5182	.5178	.5173	.5168
.64	.5163	.5159	.5154	.5150	.5145	.5141	.5137	.5132	.5128	.5124
.65	.5120	.5116	.5112	.5108	.5105	.5101	.5097	.5093	.5090	.5086
.66	.5083	.5080	.5076	.5073	.5070	.5067	.5064	.5060	.5058	.5055
.67	.5052	.5050	.5047	.5044	.5042	.5039	.5037	.5035	.5033	.5030

TABLE 1.—Continued.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.68	.5028	.5026	.5024	.5022	.5021	.5019	.5017	.5016	.5014	.5013
.69	.5011	.5010	.5009	.5008	.5007	.5006	.5005	.5004	.5003	.5003
.70	.5002	.5001	.5001	.5001	.5000	.5000	.5000	.5000	.5000	.5000
.71	.5000	.5001	.5001	.5001	.5002	.5003	.5003	.5004	.5005	.5006
.72	.5007	.5008	.5009	.5010	.5012	.5013	.5014	.5016	.5018	.5020
.73	.5022	.5024	.5026	.5028	.5030	.5032	.5035	.5037	.5040	.5043
.74	.5045	.5048	.5051	.5054	.5057	.5061	.5064	.5067	.5071	.5074
.75	.5078	.5082	.5086	.5090	.5094	.5098	.5102	.5107	.5111	.5116
.76	.5120	.5125	.5130	.5135	.5140	.5145	.5150	.5156	.5161	.5167
.77	.5173	.5178	.5184	.5190	.5196	.5203	.5209	.5215	.5222	.5228
.78	.5235	.5242	.5249	.5256	.5263	.5270	.5278	.5285	.5293	.5300
.79	.5308	.5316	.5324	.5332	.5340	.5349	.5357	.5366	.5374	.5383
.80	.5392	.5401	.5410	.5419	.5429	.5438	.5448	.5458	.5468	.5477
.81	.5487	.5498	.5508	.5518	.5529	.5539	.5550	.5561	.5572	.5583
.82	.5594	.5606	.5617	.5629	.5641	.5653	.5664	.5677	.5689	.5701
.83	.5714	.5726	.5739	.5752	.5765	.5778	.5791	.5805	.5818	.5832
.84	.5845	.5859	.5873	.5887	.5902	.5916	.5931	.5945	.5960	.5975
.85	.5990	.6005	.6021	.6036	.6052	.6067	.6083	.6099	.6115	.6132
.86	.6148	.6165	.6181	.6198	.6215	.6232	.6250	.6267	.6284	.6302
.87	.6320	.6338	.6356	.6374	.6393	.6411	.6430	.6449	.6468	.6487
.88	.6506	.6525	.6545	.6564	.6584	.6604	.6624	.6645	.6665	.6686
.89	.6706	.6727	.6748	.6770	.6791	.6812	.6834	.6856	.6878	.6900
.90	.6922	.6944	.6967	.6990	.7013	.7036	.7059	.7082	.7106	.7129
.91	.7153	.7177	.7201	.7225	.7250	.7274	.7299	.7324	.7349	.7374
.92	.7400	.7425	.7451	.7477	.7503	.7529	.7556	.7582	.7609	.7636
.93	.7663	.7690	.7718	.7745	.7773	.7801	.7829	.7857	.7886	.7914
.94	.7943	.7972	.8001	.8030	.8060	.8090	.8119	.8149	.8179	.8209
.95	.8240	.8271	.8302	.8333	.8364	.8395	.8427	.8459	.8490	.8523
.96	.8555	.8587	.8620	.8653	.8686	.8719	.8752	.8786	.8819	.8854
.97	.8888	.8922	.8957	.8991	.9026	.9061	.9096	.9132	.9168	.9203
.98	.9239	.9276	.9312	.9348	.9385	.9422	.9459	.9497	.9534	.9572
.99	.9610	.9648	.9686	.9725	.9764	.9802	.9842	.9881	.9920	.9960

tion. Thus, if the original photographs are correctly timed according to (5), the apparent Lorentz corrections required for all levels are the same, and all computations take the form

$$I \sim p \csc T |F|^2. \quad (7)$$

From this discussion, it is evident that provided time cancellation exposures are made, the Lorentz factor correction can be easily made with the aid of tables of $\csc T$ for the argument T . The angle T is usually found by measuring the location of the diffraction spot on the film in millimeters. For standard diameter cameras this linear measurement is readily converted by inspection to degrees and tenths. The argument for $\csc T$ should thus be the angle T expressed in degrees and tenths. Such tables are already extant,⁸ but for the sake of providing the reader with complete equipment for Lorentz and polarization correction, we include this information as Table III.

⁸Herbert Bristol Dwight, *Mathematical Tables* (McGraw-Hill Book Company, Inc., New York, 1941), pages 40–59.

For zero levels, both Lorentz and polarization factors can be given as functions of the same argument. We therefore provide the combined Lorentz-polarization factor, $L \cdot p$, for the arguments $\sin \theta$ and σ , in Tables IV and V, respectively. *These last two tables should be used for zero-level photographs only.*

COMPUTATIONAL FORM AND FACTORS FOR TRIAL AND ERROR INVESTIGATIONS

The computation form given in (7) contains an operation which can often be eliminated with a consequent great saving of time and labor. This is especially desirable when more than one computation of the intensity of the same reflection must be made, particularly when it must be monotonously repeated as in the case of the location of atoms by the method of trial and error. This is the operation of squaring $|F|$. If the square roots of both sides of (7) are taken, there results

$$(I)^{1/2} \sim (p \csc T)^{1/2} |F|. \quad (8)$$

TABLE II. Polarization factor, $p \left(= \frac{1 + \cos^2 2\theta}{2} \right)$, as a function of σ .

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.00	1.0000	1.0000	1.0000	1.0000	1.0000	.68	.7955	.7944	.7934	.7924	.7913
.01	.9999	.9999	.9999	.9999	.9998	.69	.7903	.7892	.7882	.7871	.7861
.02	.9998	.9998	.9997	.9997	.9996	.70	.7850	.7840	.7829	.7818	.7808
.03	.9996	.9995	.9994	.9994	.9993	.71	.7797	.7787	.7776	.7765	.7754
.04	.9992	.9991	.9990	.9989	.9989	.72	.7744	.7733	.7723	.7712	.7701
.05	.9988	.9987	.9985	.9984	.9983	.73	.7690	.7680	.7669	.7658	.7648
.06	.9982	.9981	.9980	.9978	.9977	.74	.7637	.7626	.7615	.7605	.7594
.07	.9976	.9974	.9973	.9971	.9970	.75	.7583	.7572	.7561	.7551	.7540
.08	.9968	.9966	.9965	.9963	.9961	.76	.7529	.7518	.7507	.7497	.7486
.09	.9960	.9958	.9956	.9954	.9952	.77	.7475	.7464	.7453	.7442	.7432
.10	.9950	.9948	.9946	.9944	.9942	.78	.7421	.7410	.7399	.7388	.7377
.11	.9940	.9937	.9935	.9933	.9931	.79	.7366	.7356	.7345	.7334	.7323
.12	.9928	.9926	.9923	.9921	.9918	.80	.7312	.7301	.7290	.7279	.7268
.13	.9916	.9913	.9911	.9908	.9905	.81	.7258	.7247	.7236	.7225	.7214
.14	.9902	.9900	.9897	.9894	.9891	.82	.7203	.7192	.7182	.7170	.7160
.15	.9888	.9885	.9882	.9879	.9876	.83	.7149	.7138	.7127	.7116	.7105
.16	.9873	.9870	.9866	.9863	.9860	.84	.7094	.7083	.7073	.7062	.7051
.17	.9857	.9853	.9850	.9846	.9843	.85	.7040	.7029	.7018	.7007	.6997
.18	.9839	.9836	.9832	.9828	.9825	.86	.6986	.6975	.6964	.6953	.6942
.19	.9821	.9817	.9814	.9810	.9806	.87	.6932	.6921	.6910	.6899	.6888
.20	.9802	.9798	.9794	.9790	.9786	.88	.6878	.6867	.6856	.6845	.6835
.21	.9782	.9778	.9774	.9769	.9765	.89	.6824	.6813	.6802	.6792	.6781
.22	.9761	.9757	.9752	.9748	.9743	.90	.6770	.6759	.6749	.6738	.6727
.23	.9739	.9735	.9730	.9725	.9721	.91	.6717	.6706	.6695	.6685	.6674
.24	.9716	.9711	.9707	.9702	.9697	.92	.6663	.6653	.6642	.6632	.6621
.25	.9692	.9688	.9683	.9678	.9673	.93	.6611	.6600	.6589	.6579	.6568
.26	.9668	.9663	.9658	.9652	.9647	.94	.6558	.6547	.6537	.6527	.6516
.27	.9642	.9637	.9632	.9626	.9621	.95	.6506	.6495	.6485	.6474	.6464
.28	.9616	.9610	.9605	.9599	.9594	.96	.6454	.6443	.6433	.6423	.6412
.29	.9588	.9583	.9577	.9571	.9566	.97	.6402	.6392	.6382	.6371	.6361
.30	.9560	.9554	.9549	.9543	.9537	.98	.6351	.6341	.6331	.6320	.6310
.31	.9531	.9525	.9519	.9513	.9507	.99	.6300	.6290	.6280	.6270	.6260
.32	.9501	.9495	.9489	.9483	.9477	1.00	.6250	.6240	.6230	.6220	.6210
.33	.9470	.9464	.9458	.9451	.9445	1.01	.6200	.6190	.6180	.6170	.6161
.34	.9439	.9432	.9426	.9419	.9413	1.02	.6151	.6141	.6132	.6122	.6112
.35	.9406	.9400	.9393	.9386	.9380	1.03	.6102	.6093	.6083	.6073	.6064
.36	.9373	.9366	.9360	.9353	.9346	1.04	.6054	.6044	.6035	.6026	.6016
.37	.9339	.9332	.9325	.9318	.9311	1.05	.6006	.5997	.5988	.5979	.5969
.38	.9304	.9297	.9290	.9283	.9276	1.06	.5960	.5951	.5942	.5932	.5923
.39	.9269	.9261	.9254	.9247	.9239	1.07	.5914	.5905	.5896	.5887	.5878
.40	.9232	.9225	.9217	.9210	.9202	1.08	.5869	.5860	.5851	.5842	.5833
.41	.9195	.9187	.9180	.9172	.9164	1.09	.5824	.5815	.5806	.5798	.5789
.42	.9157	.9149	.9142	.9134	.9126	1.10	.5780	.5771	.5763	.5754	.5746
.43	.9118	.9110	.9102	.9095	.9087	1.11	.5737	.5729	.5720	.5712	.5703
.44	.9079	.9071	.9063	.9055	.9047	1.12	.5695	.5687	.5678	.5670	.5662
.45	.9039	.9031	.9023	.9014	.9006	1.13	.5654	.5645	.5637	.5629	.5621
.46	.8998	.8990	.8981	.8973	.8965	1.14	.5613	.5605	.5597	.5589	.5582
.47	.8956	.8948	.8940	.8931	.8923	1.15	.5574	.5566	.5558	.5551	.5543
.48	.8914	.8906	.8897	.8888	.8880	1.16	.5535	.5528	.5520	.5513	.5505
.49	.8872	.8863	.8854	.8846	.8837	1.17	.5498	.5490	.5483	.5476	.5469
.50	.8828	.8819	.8811	.8802	.8793	1.18	.5461	.5454	.5447	.5440	.5433
.51	.8784	.8775	.8766	.8757	.8748	1.19	.5426	.5419	.5412	.5406	.5399
.52	.8739	.8730	.8721	.8712	.8703	1.20	.5392	.5385	.5379	.5372	.5365
.53	.8694	.8685	.8676	.8667	.8658	1.21	.5359	.5353	.5346	.5340	.5333
.54	.8648	.8639	.8630	.8621	.8611	1.22	.5327	.5321	.5315	.5309	.5303
.55	.8602	.8593	.8583	.8574	.8564	1.23	.5297	.5291	.5285	.5279	.5273
.56	.8555	.8545	.8536	.8526	.8517	1.24	.5267	.5262	.5256	.5250	.5245
.57	.8507	.8498	.8488	.8479	.8469	1.25	.5239	.5234	.5228	.5223	.5218
.58	.8459	.8450	.8440	.8430	.8420	1.26	.5213	.5207	.5202	.5197	.5192
.59	.8411	.8401	.8391	.8382	.8372	1.27	.5187	.5182	.5178	.5173	.5168
.60	.8362	.8352	.8342	.8332	.8323	1.28	.5163	.5159	.5154	.5150	.5145
.61	.8313	.8303	.8293	.8283	.8273	1.29	.5141	.5137	.5132	.5128	.5124
.62	.8263	.8253	.8243	.8233	.8223	1.30	.5120	.5116	.5112	.5108	.5105
.63	.8212	.8202	.8192	.8182	.8172	1.31	.5101	.5097	.5093	.5090	.5086
.64	.8162	.8152	.8141	.8131	.8121	1.32	.5083	.5080	.5076	.5073	.5070
.65	.8111	.8100	.8090	.8080	.8070	1.33	.5067	.5064	.5060	.5058	.5055
.66	.8059	.8049	.8039	.8028	.8018	1.34	.5052	.5050	.5047	.5044	.5042
.67	.8007	.7997	.7987	.7976	.7966	1.35	.5039	.5037	.5035	.5033	.5030

TABLE II.—Continued.

	.000	.002	.004	.006	.008
1.36	.5028	.5026	.5024	.5022	.5021
1.37	.5019	.5017	.5016	.5014	.5013
1.38	.5011	.5010	.5009	.5008	.5007
1.39	.5006	.5005	.5004	.5003	.5003
1.40	.5002	.5001	.5001	.5001	.5000
1.41	.5000	.5000	.5000	.5000	.5000
1.42	.5000	.5001	.5001	.5001	.5002
1.43	.5003	.5003	.5004	.5005	.5006
1.44	.5007	.5008	.5009	.5010	.5012
1.45	.5013	.5014	.5016	.5018	.5020
1.46	.5022	.5024	.5026	.5028	.5030
1.47	.5032	.5035	.5037	.5040	.5043
1.48	.5045	.5048	.5051	.5054	.5057
1.49	.5061	.5064	.5067	.5071	.5074
1.50	.5078	.5082	.5086	.5090	.5094
1.51	.5098	.5102	.5107	.5111	.5116
1.52	.5120	.5125	.5130	.5135	.5140
1.53	.5145	.5150	.5156	.5161	.5167
1.54	.5173	.5178	.5184	.5190	.5196
1.55	.5203	.5209	.5215	.5222	.5228
1.56	.5235	.5242	.5249	.5256	.5263
1.57	.5270	.5278	.5285	.5293	.5300
1.58	.5308	.5316	.5324	.5332	.5340
1.59	.5349	.5357	.5366	.5374	.5383
1.60	.5392	.5401	.5410	.5419	.5429
1.61	.5438	.5448	.5458	.5468	.5477
1.62	.5487	.5498	.5508	.5518	.5529
1.63	.5539	.5550	.5561	.5572	.5583
1.64	.5594	.5606	.5617	.5629	.5641
1.65	.5653	.5664	.5677	.5689	.5701
1.66	.5714	.5726	.5739	.5752	.5765
1.67	.5778	.5791	.5805	.5818	.5832
1.68	.5845	.5859	.5873	.5887	.5902
1.69	.5916	.5931	.5945	.5960	.5975
1.70	.5990	.6005	.6021	.6036	.6052
1.71	.6067	.6083	.6099	.6115	.6132
1.72	.6148	.6165	.6181	.6198	.6215
1.73	.6232	.6250	.6267	.6284	.6302
1.74	.6320	.6338	.6356	.6374	.6393
1.75	.6411	.6430	.6449	.6468	.6487
1.76	.6506	.6525	.6545	.6564	.6584
1.77	.6604	.6624	.6645	.6665	.6686
1.78	.6706	.6727	.6748	.6770	.6791
1.79	.6812	.6834	.6856	.6878	.6900
1.80	.6922	.6944	.6967	.6990	.7013
1.81	.7036	.7059	.7082	.7106	.7129
1.82	.7153	.7177	.7201	.7225	.7250
1.83	.7274	.7299	.7324	.7349	.7374
1.84	.7400	.7425	.7451	.7477	.7503
1.85	.7529	.7556	.7582	.7609	.7636
1.86	.7663	.7690	.7718	.7745	.7773
1.87	.7801	.7829	.7857	.7886	.7914
1.88	.7943	.7972	.8001	.8030	.8060
1.89	.8090	.8119	.8149	.8179	.8209
1.90	.8240	.8271	.8302	.8333	.8364
1.91	.8395	.8427	.8459	.8490	.8523
1.92	.8555	.8587	.8620	.8653	.8686
1.93	.8719	.8752	.8786	.8819	.8854
1.94	.8888	.8922	.8957	.8991	.9026
1.95	.9061	.9096	.9132	.9168	.9203
1.96	.9239	.9276	.9312	.9348	.9385
1.97	.9422	.9459	.9497	.9534	.9572
1.98	.9610	.9648	.9686	.9725	.9764
1.99	.9802	.9842	.9881	.9920	.9960

Advantage may be taken of this shortened computation in the event that the experimental intensities are not measured on a quantitative

scale, but are merely compared by means of inequalities. If the several spectra hkl are arranged in order of decreasing intensity, I , then they are also arranged in order of decreasing $(I)^{1/2}$. Thus, if an arrangement of atoms can be found whose computed $(I)^{1/2}$'s are in the same order as the observed blackening on the x-ray photographs, the computed I 's are in the same order. In order to make use of this shortened form for trial and error computations, we present Tables VI–X, which are the square roots of the functions given in Tables I–V, respectively.

There are further advantages⁴ to the use of square-root correction factors. This can be realized by arranging the computations in a certain form. This is derived as follows:

The amplitude, F , is complex, and, for computational purposes, is split into a real component and an imaginary component. Calling these components A and B for atoms of unit scattering power, the scattering power of a collection of actual atoms is given by

$$|F|^2 = (f_1A_1 + f_2A_2 + \dots)^2 + (f_1B_1 + f_2B_2 + \dots)^2, \quad (9)$$

where the subscripts 1, 2, $\dots$ relate to the first, second, $\dots$, etc., atoms in the cell, and f is the scattering power of the particular atom for that reflection. Combining this with (7), the computation becomes

$$\begin{aligned} I \sim p \csc T \{ & (f_1A_1 + f_2A_2 + \dots)^2 \\ & + (f_1B_1 + f_2B_2 + \dots)^2 \} \\ \sim \{ & (p \csc T)^{1/2} (f_1A_1 + f_2A_2 + \dots) \}^2 \\ & + \{ (p \csc T)^{1/2} (f_1B_1 + f_2B_2 + \dots) \}^2. \end{aligned} \quad (10)$$

It is convenient and usual to lump together the fA (and also the fB) terms for sets of identical atoms which are related by space group operations so that the terms in the parentheses of (9) are condensed to produce a "structure factor." This condensation can be symbolically represented by

$$f_1A_1 + f_1A_2 + f_1A_3 + \dots = f_1m_1\bar{A}_1. \quad (11)$$

Here, $m_1\bar{A}_1$ is the "structure factor" of the atom set I. It is composed of an integral part, m_1 , which is the number of identical atoms in the set, and a trigonometric part $\bar{A}_1$. The form of the

⁴ M. J. Buerger, "Numerical structure factor tables," Special Paper Number 33, Geological Society of America (1941), pages 7–11.

TABLE III. Unscaled part of Lorentz factor, $S \cdot L (= \csc T)$, as a function of T .

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	∞	572.96	286.48	190.99	143.24	114.59	95.495	81.853	71.622	63.665
1	57.299	52.090	47.750	44.077	40.930	38.202	35.815	33.708	31.836	30.161
2	28.654	27.290	26.050	24.918	23.880	22.926	22.044	21.229	20.471	19.766
3	19.107	18.492	17.914	17.372	16.862	16.380	15.926	15.496	15.089	14.703
4	14.336	13.987	13.654	13.337	13.035	12.745	12.469	12.204	11.951	11.707
5	11.474	11.249	11.034	10.826	10.626	10.433	10.248	10.068	9.895	9.728
6	9.567	9.411	9.259	9.113	8.971	8.834	8.700	8.571	8.446	8.324
7	8.206	8.091	7.979	7.870	7.764	7.661	7.561	7.464	7.368	7.276
8	7.185	7.097	7.011	6.927	6.845	6.766	6.687	6.611	6.537	6.464
9	6.393	6.323	6.255	6.188	6.123	6.059	5.996	5.935	5.875	5.816
10	5.759	5.702	5.647	5.593	5.540	5.487	5.436	5.386	5.337	5.288
11	5.241	5.194	5.148	5.103	5.059	5.016	4.973	4.931	4.890	4.850
12	4.810	4.771	4.732	4.694	4.657	4.620	4.584	4.549	4.514	4.479
13	4.445	4.412	4.379	4.347	4.315	4.284	4.253	4.222	4.192	4.163
14	4.134	4.105	4.077	4.049	4.021	3.994	3.967	3.941	3.915	3.889
15	3.864	3.839	3.814	3.790	3.766	3.742	3.719	3.696	3.673	3.650
16	3.628	3.606	3.584	3.563	3.542	3.521	3.500	3.480	3.460	3.440
17	3.420	3.401	3.382	3.363	3.344	3.326	3.307	3.289	3.271	3.254
18	3.236	3.219	3.202	3.185	3.168	3.152	3.135	3.119	3.103	3.087
19	3.072	3.056	3.041	3.026	3.011	2.996	2.981	2.967	2.952	2.938
20	2.924	2.910	2.896	2.882	2.869	2.856	2.842	2.829	2.816	2.803
21	2.790	2.778	2.765	2.753	2.741	2.729	2.717	2.705	2.693	2.681
22	2.670	2.658	2.647	2.635	2.624	2.613	2.602	2.591	2.581	2.570
23	2.559	2.549	2.538	2.528	2.518	2.508	2.498	2.488	2.478	2.468
24	2.459	2.449	2.440	2.430	2.421	2.411	2.402	2.393	2.384	2.375
25	2.366	2.357	2.349	2.340	2.331	2.323	2.314	2.306	2.298	2.289
26	2.281	2.273	2.265	2.257	2.249	2.241	2.233	2.226	2.218	2.210
27	2.203	2.195	2.188	2.180	2.173	2.166	2.158	2.151	2.144	2.137
28	2.130	2.123	2.116	2.109	2.103	2.096	2.089	2.082	2.076	2.069
29	2.063	2.056	2.050	2.043	2.037	2.031	2.025	2.018	2.012	2.006
30	2.000	1.994	1.988	1.982	1.976	1.970	1.964	1.959	1.953	1.947
31	1.942	1.936	1.930	1.925	1.919	1.914	1.908	1.903	1.898	1.892
32	1.887	1.882	1.877	1.871	1.866	1.861	1.856	1.851	1.846	1.841
33	1.836	1.831	1.826	1.821	1.817	1.812	1.807	1.802	1.798	1.793
34	1.788	1.784	1.779	1.775	1.770	1.766	1.761	1.757	1.752	1.748
35	1.743	1.739	1.735	1.731	1.726	1.722	1.718	1.714	1.710	1.705
36	1.701	1.697	1.693	1.689	1.685	1.681	1.677	1.673	1.669	1.666
37	1.662	1.658	1.654	1.650	1.646	1.643	1.639	1.635	1.632	1.628
38	1.624	1.621	1.617	1.613	1.610	1.606	1.603	1.599	1.596	1.592
39	1.589	1.586	1.582	1.579	1.575	1.572	1.569	1.566	1.562	1.559
40	1.556	1.553	1.549	1.546	1.543	1.540	1.537	1.534	1.530	1.527
41	1.524	1.521	1.518	1.515	1.512	1.509	1.506	1.503	1.500	1.497
42	1.494	1.492	1.489	1.486	1.483	1.480	1.477	1.475	1.472	1.469
43	1.466	1.464	1.461	1.458	1.455	1.453	1.450	1.447	1.445	1.442
44	1.440	1.437	1.434	1.432	1.429	1.427	1.424	1.422	1.419	1.417
45	1.414	1.412	1.409	1.407	1.404	1.402	1.400	1.397	1.395	1.393
46	1.390	1.388	1.386	1.383	1.381	1.379	1.376	1.374	1.372	1.370
47	1.367	1.365	1.363	1.361	1.359	1.356	1.354	1.352	1.350	1.348
48	1.346	1.344	1.341	1.339	1.337	1.335	1.333	1.331	1.329	1.327
49	1.325	1.323	1.321	1.319	1.317	1.315	1.313	1.311	1.309	1.307
50	1.305	1.304	1.302	1.300	1.298	1.296	1.294	1.292	1.290	1.289
51	1.287	1.285	1.283	1.281	1.280	1.278	1.276	1.274	1.273	1.271
52	1.269	1.267	1.266	1.264	1.262	1.260	1.259	1.257	1.255	1.254
53	1.252	1.250	1.249	1.247	1.246	1.244	1.242	1.241	1.239	1.238
54	1.236	1.235	1.233	1.231	1.230	1.228	1.227	1.225	1.224	1.222
55	1.221	1.219	1.218	1.216	1.215	1.213	1.212	1.211	1.209	1.208
56	1.206	1.205	1.203	1.202	1.201	1.199	1.198	1.196	1.195	1.194
57	1.192	1.191	1.190	1.188	1.187	1.186	1.184	1.183	1.182	1.180
58	1.179	1.178	1.177	1.175	1.174	1.173	1.172	1.170	1.169	1.168
59	1.167	1.165	1.164	1.163	1.162	1.161	1.159	1.158	1.157	1.156
60	1.155	1.154	1.152	1.151	1.150	1.149	1.148	1.147	1.146	1.144
61	1.143	1.142	1.141	1.140	1.139	1.138	1.137	1.136	1.135	1.134
62	1.133	1.132	1.130	1.129	1.128	1.127	1.126	1.125	1.124	1.123
63	1.122	1.121	1.120	1.119	1.118	1.117	1.116	1.115	1.115	1.114
64	1.113	1.112	1.111	1.110	1.109	1.108	1.107	1.106	1.105	1.104
65	1.103	1.102	1.102	1.101	1.100	1.099	1.098	1.097	1.096	1.095
66	1.095	1.094	1.093	1.092	1.091	1.090	1.090	1.089	1.088	1.087
67	1.086	1.086	1.085	1.084	1.083	1.082	1.082	1.081	1.080	1.079
68	1.079	1.078	1.077	1.076	1.076	1.075	1.074	1.073	1.073	1.072
69	1.071	1.070	1.070	1.069	1.068	1.068	1.067	1.066	1.066	1.065

TABLE III.—Continued.

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
70	1.064	1.064	1.063	1.062	1.062	1.061	1.060	1.060	1.059	1.058
71	1.058	1.057	1.056	1.056	1.055	1.054	1.054	1.053	1.053	1.052
72	1.051	1.051	1.050	1.050	1.049	1.049	1.048	1.047	1.047	1.046
73	1.046	1.045	1.045	1.044	1.043	1.043	1.042	1.042	1.041	1.041
74	1.040	1.040	1.039	1.039	1.038	1.038	1.037	1.037	1.036	1.036
75	1.035	1.035	1.034	1.034	1.033	1.033	1.032	1.032	1.032	1.031
76	1.031	1.030	1.030	1.029	1.029	1.028	1.028	1.028	1.027	1.027
77	1.026	1.026	1.025	1.025	1.025	1.024	1.024	1.023	1.023	1.023
78	1.022	1.022	1.022	1.021	1.021	1.020	1.020	1.020	1.019	1.019
79	1.019	1.018	1.018	1.018	1.017	1.017	1.017	1.016	1.016	1.016
80	1.015	1.015	1.015	1.015	1.014	1.014	1.014	1.013	1.013	1.013
81	1.012	1.012	1.012	1.012	1.011	1.011	1.011	1.011	1.010	1.010
82	1.010	1.010	1.009	1.009	1.009	1.009	1.008	1.008	1.008	1.008
83	1.008	1.007	1.007	1.007	1.007	1.006	1.006	1.006	1.006	1.006
84	1.006	1.005	1.005	1.005	1.005	1.005	1.004	1.004	1.004	1.004
85	1.004	1.004	1.004	1.003	1.003	1.003	1.003	1.003	1.003	1.003
86	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.001
87	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
88	1.001	1.001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
89	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
91	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.001
92	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
93	1.001	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002
94	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.004	1.004	1.004
95	1.004	1.004	1.004	1.004	1.005	1.005	1.005	1.005	1.005	1.006
96	1.006	1.006	1.006	1.006	1.006	1.007	1.007	1.007	1.007	1.008
97	1.008	1.008	1.008	1.008	1.009	1.009	1.009	1.009	1.010	1.010
98	1.010	1.010	1.011	1.011	1.011	1.011	1.012	1.012	1.012	1.012
99	1.013	1.013	1.013	1.014	1.014	1.014	1.015	1.015	1.015	1.015
100	1.016	1.016	1.016	1.017	1.017	1.017	1.018	1.018	1.018	1.019
101	1.019	1.019	1.020	1.020	1.020	1.021	1.021	1.022	1.022	1.022
102	1.023	1.023	1.023	1.024	1.024	1.025	1.025	1.025	1.026	1.026
103	1.027	1.027	1.028	1.028	1.028	1.029	1.029	1.030	1.030	1.031
104	1.031	1.032	1.032	1.032	1.033	1.033	1.034	1.034	1.035	1.035
105	1.036	1.036	1.037	1.037	1.038	1.038	1.039	1.039	1.040	1.040
106	1.041	1.041	1.042	1.042	1.043	1.043	1.044	1.045	1.045	1.046
107	1.046	1.047	1.047	1.048	1.049	1.049	1.050	1.050	1.051	1.051
108	1.052	1.053	1.053	1.054	1.054	1.055	1.056	1.056	1.057	1.058
109	1.058	1.059	1.060	1.060	1.061	1.062	1.062	1.063	1.064	1.064
110	1.065	1.066	1.066	1.067	1.068	1.068	1.069	1.070	1.070	1.071
111	1.072	1.073	1.073	1.074	1.075	1.076	1.076	1.077	1.078	1.079
112	1.079	1.080	1.081	1.082	1.082	1.083	1.084	1.085	1.086	1.086
113	1.087	1.088	1.089	1.090	1.090	1.091	1.092	1.093	1.094	1.095
114	1.095	1.096	1.097	1.098	1.099	1.100	1.101	1.102	1.102	1.103
115	1.104	1.105	1.106	1.107	1.108	1.109	1.110	1.111	1.112	1.113
116	1.114	1.115	1.115	1.116	1.117	1.118	1.119	1.120	1.121	1.122
117	1.123	1.124	1.125	1.126	1.127	1.128	1.129	1.130	1.132	1.133
118	1.134	1.135	1.136	1.137	1.138	1.139	1.140	1.141	1.142	1.143
119	1.144	1.146	1.147	1.148	1.149	1.150	1.151	1.152	1.154	1.155
120	1.156	1.157	1.158	1.159	1.161	1.162	1.163	1.164	1.165	1.167
121	1.168	1.169	1.170	1.172	1.173	1.174	1.175	1.177	1.178	1.179
122	1.180	1.182	1.183	1.184	1.186	1.187	1.188	1.190	1.191	1.192
123	1.194	1.195	1.196	1.198	1.199	1.201	1.202	1.203	1.205	1.206
124	1.208	1.209	1.211	1.212	1.213	1.215	1.216	1.218	1.219	1.221
125	1.222	1.224	1.225	1.227	1.228	1.230	1.231	1.233	1.235	1.236
126	1.238	1.239	1.241	1.242	1.244	1.246	1.247	1.249	1.250	1.252
127	1.254	1.255	1.257	1.259	1.260	1.262	1.264	1.266	1.267	1.269
128	1.271	1.273	1.274	1.276	1.278	1.280	1.281	1.283	1.285	1.287
129	1.289	1.290	1.292	1.294	1.296	1.298	1.300	1.302	1.304	1.305
130	1.307	1.309	1.311	1.313	1.315	1.317	1.319	1.321	1.323	1.325
131	1.327	1.329	1.331	1.333	1.335	1.337	1.339	1.341	1.344	1.346
132	1.348	1.350	1.352	1.354	1.356	1.359	1.361	1.363	1.365	1.367
133	1.370	1.372	1.374	1.376	1.379	1.381	1.383	1.386	1.388	1.390
134	1.393	1.395	1.397	1.400	1.402	1.404	1.407	1.409	1.412	1.414
135	1.417	1.419	1.422	1.424	1.427	1.429	1.432	1.434	1.437	1.440
136	1.442	1.445	1.447	1.450	1.453	1.455	1.458	1.461	1.464	1.466
137	1.469	1.472	1.475	1.477	1.480	1.483	1.486	1.489	1.492	1.494
138	1.497	1.500	1.503	1.506	1.509	1.512	1.515	1.518	1.521	1.524
139	1.527	1.530	1.534	1.537	1.540	1.543	1.546	1.549	1.553	1.556

TABLE III.—Continued.

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
140	1.559	1.562	1.566	1.569	1.572	1.575	1.579	1.582	1.586	1.589
141	1.592	1.596	1.599	1.603	1.606	1.610	1.613	1.617	1.621	1.624
142	1.628	1.632	1.635	1.639	1.643	1.646	1.650	1.654	1.658	1.662
143	1.666	1.669	1.673	1.677	1.681	1.685	1.689	1.693	1.697	1.701
144	1.705	1.710	1.714	1.718	1.722	1.726	1.731	1.735	1.739	1.743
145	1.748	1.752	1.757	1.761	1.766	1.770	1.775	1.779	1.784	1.788
146	1.793	1.798	1.802	1.807	1.812	1.817	1.821	1.826	1.831	1.836
147	1.841	1.846	1.851	1.856	1.861	1.866	1.871	1.877	1.882	1.887
148	1.892	1.898	1.903	1.908	1.914	1.919	1.925	1.930	1.936	1.942
149	1.947	1.953	1.959	1.964	1.970	1.976	1.982	1.988	1.994	2.000
150	2.006	2.012	2.018	2.025	2.031	2.037	2.043	2.050	2.056	2.063
151	2.069	2.076	2.082	2.089	2.096	2.103	2.109	2.116	2.123	2.130
152	2.137	2.144	2.151	2.158	2.166	2.173	2.180	2.188	2.195	2.203
153	2.210	2.218	2.226	2.233	2.241	2.249	2.257	2.265	2.273	2.281
154	2.289	2.298	2.306	2.314	2.323	2.331	2.340	2.349	2.357	2.366
155	2.375	2.384	2.393	2.402	2.411	2.421	2.430	2.440	2.449	2.459
156	2.468	2.478	2.488	2.498	2.508	2.518	2.528	2.538	2.549	2.559
157	2.570	2.581	2.591	2.602	2.613	2.624	2.635	2.647	2.658	2.670
158	2.681	2.693	2.705	2.717	2.729	2.741	2.753	2.765	2.778	2.790
159	2.803	2.816	2.829	2.842	2.856	2.869	2.882	2.896	2.910	2.924
160	2.938	2.952	2.967	2.981	2.996	3.011	3.026	3.041	3.056	3.072
161	3.087	3.103	3.119	3.135	3.152	3.168	3.185	3.202	3.219	3.236
162	3.254	3.271	3.289	3.307	3.326	3.344	3.363	3.382	3.401	3.420
163	3.440	3.460	3.480	3.500	3.521	3.542	3.563	3.584	3.606	3.628
164	3.650	3.673	3.696	3.719	3.742	3.766	3.790	3.814	3.839	3.864
165	3.889	3.915	3.941	3.967	3.994	4.021	4.049	4.077	4.105	4.134
166	4.163	4.192	4.222	4.253	4.284	4.315	4.347	4.379	4.412	4.445
167	4.479	4.514	4.549	4.584	4.620	4.657	4.694	4.732	4.771	4.810
168	4.850	4.890	4.931	4.973	5.016	5.059	5.103	5.148	5.194	5.241
169	5.288	5.337	5.386	5.436	5.487	5.540	5.593	5.647	5.702	5.759
170	5.816	5.875	5.935	5.996	6.059	6.123	6.188	6.255	6.323	6.393
171	6.464	6.537	6.611	6.687	6.766	6.845	6.927	7.011	7.097	7.185
172	7.276	7.368	7.464	7.561	7.661	7.764	7.870	7.979	8.091	8.206
173	8.324	8.446	8.571	8.700	8.834	8.971	9.113	9.259	9.411	9.567
174	9.728	9.895	10.068	10.248	10.433	10.626	10.826	11.034	11.249	11.474
175	11.707	11.951	12.204	12.469	12.745	13.035	13.337	13.654	13.987	14.336
176	14.703	15.089	15.496	15.926	16.380	16.862	17.372	17.914	18.492	19.107
177	19.766	20.471	21.229	22.044	22.926	23.880	24.918	26.050	27.290	28.654
178	30.161	31.836	33.708	35.815	38.202	40.930	44.077	47.750	52.090	57.299
179	63.665	71.622	81.853	95.495	114.59	143.24	190.99	286.48	572.96	∞

entire structure factor is available in several places⁶ and the variation of the trigonometric part with the location of a representative atom of the set is also available.^{4,7} Substituting (11) in (10), it becomes

$$I \sim \{[(p \csc T)^{\frac{1}{2}} f_I m_I] \bar{A}_I + [(p \csc T)^{\frac{1}{2}} f_{II} m_{II}] \bar{A}_{II} + \dots\}^2 + \{[(p \csc T)^{\frac{1}{2}} f_I m_I] \bar{B}_I + [(p \csc T)^{\frac{1}{2}} f_{II} m_{II}] \bar{B}_{II} + \dots\}^2. \quad (13)$$

The advantage in using the square roots of the corrections is that the computations repre-

⁶ *International Tables for the Determination of Crystal Structures* (Gebrüder Borntraeger, Berlin, 1935).

⁷ Kathleen Lonsdale, *Simplified Structure Factor and Electron Density Formulae for the 230 Space Groups of Mathematical Crystallography* (G. Bell & Sons, Ltd., London, 1936).

⁸ W. L. Bragg and H. Lipson, "The employment of contour graphs of structure-factor in crystal analysis," *Zeits. f. Krist. (A)* **95**, 323-337 (1936).

$$I \sim \{(p \csc T)^{\frac{1}{2}} (f_I m_I \bar{A}_I + f_{II} m_{II} \bar{A}_{II} + \dots)\}^2 + \{(p \csc T)^{\frac{1}{2}} (f_I m_I \bar{B}_I + f_{II} m_{II} \bar{B}_{II} + \dots)\}^2. \quad (12)$$

For computational purposes, this is expanded and arranged in columns as follows:

sented by the several terms in square brackets need be made only once for each particular reflection for the entire trial and error computation process. It is applied as a coefficient to the only variable parts of the computation, namely $\bar{A}$ and $\bar{B}$. Designating this coefficient by

$$C = (p \csc T)^{\frac{1}{2}} f m, \quad (14)$$

TABLE IV. Lorentz-polarization factor, $L \cdot p \left(= \frac{1}{\sin 2\theta} \cdot \frac{1 + \cos^2 2\theta}{2} \right)$, as a function of $\sin \theta$.
Valid only for zero level when $\mu = 0$.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.00	∞	502.51	249.37	166.66	125.62	100.19	83.327	71.268	62.531	55.516
.01	49.990	45.464	41.637	38.448	35.713	33.340	31.234	29.403	27.775	26.297
.02	24.986	23.800	22.705	21.721	20.818	19.975	19.208	18.501	17.842	17.218
.03	16.645	16.111	15.598	15.127	14.682	14.257	13.862	13.489	13.128	12.786
.04	12.472	12.163	11.874	11.598	11.330	11.078	10.833	10.602	10.382	10.165
.05	9.9675	9.7664	9.5755	9.3944	9.2199	9.0492	8.8870	8.7312	8.5777	8.4315
.06	8.2880	8.1513	8.0176	7.8894	7.7658	7.6434	7.5212	7.4120	7.3024	7.1938
.07	7.0904	6.9898	6.8905	6.7953	6.7013	6.6110	6.5218	6.4362	6.3514	6.2704
.08	6.1899	6.1123	6.0359	5.9621	5.8893	5.8192	5.7493	5.6823	5.6158	5.5518
.09	5.4882	5.4274	5.3669	5.3072	5.2488	5.1928	5.1367	5.0821	5.0291	4.9766
.10	4.9258	4.8751	4.8268	4.7777	4.7300	4.6841	4.6382	4.5937	4.5494	4.5061
.11	4.4642	4.4222	4.3810	4.3413	4.3013	4.2623	4.2247	4.1869	4.1497	4.1137
.12	4.0779	4.0425	4.0079	3.9743	3.9405	3.9074	3.8748	3.8434	3.8117	3.7806
.13	3.7499	3.7203	3.6906	3.6612	3.6324	3.6040	3.5764	3.5487	3.5216	3.4947
.14	3.4682	3.4422	3.4164	3.3915	3.3664	3.3417	3.3174	3.2934	3.2696	3.2488
.15	3.2231	3.2003	3.1779	3.1556	3.1337	3.1120	3.0906	3.0696	3.0487	3.0281
.16	3.0081	2.9877	2.9679	2.9482	2.9292	2.9097	2.8908	2.8721	2.8533	2.8350
.17	2.8170	2.7992	2.7815	2.7641	2.7468	2.7295	2.7126	2.6960	2.6794	2.6629
.18	2.6467	2.6308	2.6150	2.5991	2.5837	2.5684	2.5533	2.5380	2.5232	2.5086
.19	2.4938	2.4794	2.4650	2.4509	2.4370	2.4230	2.4093	2.3956	2.3821	2.3689
.20	2.3556	2.3426	2.3295	2.3167	2.3039	2.2914	2.2788	2.2663	2.2541	2.2419
.21	2.2298	2.2180	2.2060	2.1944	2.1828	2.1712	2.1599	2.1485	2.1372	2.1262
.22	2.1152	2.1042	2.0933	2.0827	2.0720	2.0614	2.0509	2.0405	2.0303	2.0201
.23	2.0099	1.9999	1.9899	1.9800	1.9702	1.9605	1.9508	1.9412	1.9317	1.9223
.24	1.9129	1.9037	1.8945	1.8854	1.8763	1.8674	1.8585	1.8497	1.8409	1.8322
.25	1.8234	1.8149	1.8064	1.7980	1.7896	1.7812	1.7729	1.7648	1.7567	1.7485
.26	1.7405	1.7325	1.7246	1.7167	1.7089	1.7012	1.6936	1.6859	1.6783	1.6707
.27	1.6633	1.6559	1.6485	1.6412	1.6339	1.6267	1.6195	1.6124	1.6053	1.5982
.28	1.5913	1.5843	1.5774	1.5707	1.5638	1.5571	1.5503	1.5438	1.5372	1.5306
.29	1.5240	1.5175	1.5110	1.5046	1.4982	1.4919	1.4856	1.4794	1.4732	1.4670
.30	1.4609	1.4548	1.4488	1.4428	1.4368	1.4309	1.4250	1.4191	1.4133	1.4075
.31	1.4017	1.3960	1.3903	1.3847	1.3791	1.3734	1.3679	1.3624	1.3569	1.3515
.32	1.3460	1.3407	1.3353	1.3300	1.3247	1.3193	1.3142	1.3089	1.3038	1.2986
.33	1.2935	1.2885	1.2834	1.2784	1.2733	1.2684	1.2635	1.2586	1.2537	1.2488
.34	1.2440	1.2392	1.2344	1.2296	1.2249	1.2202	1.2156	1.2109	1.2063	1.2017
.35	1.1972	1.1926	1.1881	1.1836	1.1791	1.1747	1.1702	1.1659	1.1615	1.1571
.36	1.1528	1.1485	1.1442	1.1399	1.1357	1.1315	1.1273	1.1232	1.1191	1.1149
.37	1.1108	1.1067	1.1027	1.0986	1.0946	1.0907	1.0867	1.0827	1.0788	1.0748
.38	1.0709	1.0671	1.0632	1.0594	1.0556	1.0518	1.0480	1.0443	1.0406	1.0369
.39	1.0331	1.0295	1.0258	1.0222	1.0185	1.0150	1.0114	1.0078	1.0043	1.0007
.40	.9972	.9937	.9903	.9868	.9834	.9800	.9765	.9731	.9698	.9664
.41	.9631	.9597	.9564	.9532	.9499	.9467	.9434	.9402	.9370	.9338
.42	.9306	.9274	.9243	.9212	.9181	.9149	.9119	.9088	.9057	.9027
.43	.8997	.8967	.8937	.8907	.8877	.8848	.8819	.8789	.8761	.8732
.44	.8703	.8674	.8646	.8618	.8590	.8561	.8534	.8506	.8478	.8450
.45	.8423	.8396	.8369	.8342	.8315	.8288	.8262	.8235	.8209	.8183
.46	.8157	.8131	.8105	.8080	.8054	.8029	.8003	.7978	.7953	.7929
.47	.7904	.7879	.7854	.7830	.7806	.7782	.7758	.7734	.7710	.7687
.48	.7663	.7639	.7616	.7593	.7570	.7547	.7524	.7501	.7479	.7456
.49	.7434	.7412	.7390	.7368	.7346	.7324	.7302	.7281	.7259	.7238
.50	.7217	.7195	.7175	.7153	.7133	.7112	.7091	.7071	.7051	.7031
.51	.7010	.6991	.6971	.6951	.6931	.6912	.6892	.6873	.6853	.6834
.52	.6815	.6796	.6777	.6758	.6740	.6722	.6703	.6685	.6667	.6648
.53	.6630	.6612	.6595	.6577	.6559	.6542	.6524	.6507	.6490	.6473
.54	.6456	.6439	.6422	.6406	.6389	.6372	.6356	.6340	.6324	.6308
.55	.6292	.6276	.6260	.6249	.6229	.6213	.6198	.6182	.6167	.6152
.56	.6137	.6122	.6107	.6092	.6078	.6064	.6049	.6035	.6021	.6006
.57	.5993	.5978	.5965	.5951	.5937	.5924	.5910	.5897	.5884	.5871
.58	.5857	.5844	.5832	.5819	.5806	.5794	.5781	.5769	.5756	.5744
.59	.5732	.5720	.5708	.5696	.5685	.5673	.5661	.5650	.5639	.5628
.60	.5616	.5605	.5594	.5583	.5573	.5562	.5552	.5541	.5531	.5520
.61	.5510	.5500	.5490	.5480	.5470	.5461	.5451	.5442	.5432	.5423
.62	.5414	.5405	.5396	.5387	.5378	.5369	.5360	.5352	.5343	.5335
.63	.5327	.5319	.5311	.5303	.5295	.5287	.5279	.5272	.5264	.5257
.64	.5250	.5243	.5235	.5229	.5222	.5215	.5208	.5202	.5195	.5189
.65	.5182	.5176	.5170	.5164	.5158	.5153	.5147	.5141	.5136	.5131
.66	.5125	.5120	.5115	.5110	.5105	.5101	.5096	.5091	.5087	.5083

TABLE IV.—Continued.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.67	.5079	.5075	.5070	.5067	.5063	.5059	.5056	.5052	.5049	.5045
.68	.5042	.5039	.5036	.5034	.5031	.5028	.5026	.5023	.5021	.5019
.69	.5017	.5015	.5013	.5011	.5010	.5008	.5007	.5006	.5005	.5004
.70	.5003	.5002	.5001	.5001	.5000	.5000	.5000	.5000	.5000	.5000
.71	.5000	.5001	.5001	.5002	.5003	.5004	.5005	.5006	.5007	.5008
.72	.5010	.5012	.5013	.5015	.5017	.5019	.5022	.5024	.5027	.5030
.73	.5032	.5035	.5038	.5042	.5045	.5048	.5052	.5056	.5060	.5064
.74	.5068	.5072	.5077	.5081	.5086	.5091	.5096	.5101	.5107	.5112
.75	.5118	.5124	.5130	.5136	.5142	.5148	.5155	.5162	.5169	.5176
.76	.5183	.5190	.5198	.5205	.5213	.5221	.5230	.5238	.5247	.5255
.77	.5264	.5273	.5282	.5292	.5301	.5311	.5321	.5331	.5341	.5352
.78	.5362	.5373	.5384	.5396	.5407	.5418	.5430	.5442	.5454	.5467
.79	.5479	.5492	.5505	.5518	.5532	.5545	.5559	.5573	.5587	.5602
.80	.5616	.5631	.5647	.5662	.5677	.5693	.5709	.5726	.5742	.5759
.81	.5776	.5793	.5811	.5828	.5846	.5864	.5883	.5902	.5921	.5940
.82	.5960	.5980	.6000	.6020	.6041	.6061	.6083	.6104	.6126	.6148
.83	.6171	.6194	.6217	.6240	.6264	.6288	.6312	.6337	.6362	.6387
.84	.6412	.6438	.6465	.6491	.6518	.6546	.6574	.6602	.6630	.6659
.85	.6688	.6718	.6749	.6779	.6810	.6841	.6873	.6905	.6938	.6971
.86	.7005	.7039	.7073	.7108	.7143	.7179	.7216	.7253	.7290	.7328
.87	.7367	.7405	.7445	.7485	.7526	.7567	.7609	.7652	.7694	.7738
.88	.7782	.7827	.7873	.7919	.7966	.8014	.8062	.8111	.8161	.8212
.89	.8263	.8315	.8368	.8422	.8476	.8531	.8588	.8645	.8703	.8762
.90	.8822	.8883	.8945	.9008	.9072	.9137	.9203	.9270	.9338	.9409
.91	.9479	.9551	.9624	.9699	.9775	.9852	.9931	1.0011	1.0093	1.0176
.92	1.0261	1.0348	1.0436	1.0526	1.0618	1.0711	1.0807	1.0904	1.1003	1.1105
.93	1.1209	1.1314	1.1423	1.1533	1.1647	1.1762	1.1881	1.2002	1.2126	1.2253
.94	1.2384	1.2517	1.2654	1.2794	1.2939	1.3086	1.3238	1.3393	1.3555	1.3720
.95	1.3889	1.4064	1.4244	1.4430	1.4621	1.4819	1.5023	1.5235	1.5452	1.5678
.96	1.5913	1.6156	1.6408	1.6670	1.6942	1.7226	1.7521	1.7831	1.8153	1.8491
.97	1.8845	1.9217	1.9608	2.0019	2.0451	2.0911	2.1399	2.1917	2.2469	2.3056
.98	2.3689	2.4367	2.5103	2.5898	2.6767	2.7719	2.8766	2.9932	3.1239	3.2717
.99	3.4403	3.6364	3.8671	4.1458	4.4898	4.9321	5.5296	6.4024	7.8633	11.152

TABLE V. Lorentz-polarization factor $F_A = \frac{1}{\sin 2\theta} \cdot \frac{1 + \cos^2 2\theta}{2}$ as a function of α .Valid only for zero level when $\mu=0$.

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.00	∞	502.51	249.37	166.66	125.62	.29	3.3417	3.3174	3.2934	3.2696	3.2488
.01	100.19	83.327	71.268	62.531	55.516	.30	3.2231	3.2003	3.1779	3.1556	3.1337
.02	49.990	45.464	41.637	38.448	35.713	.31	3.1120	3.0906	3.0696	3.0487	3.0281
.03	33.340	31.234	29.403	27.775	26.297	.32	3.0081	2.9877	2.9679	2.9482	2.9292
.04	24.986	23.800	22.705	21.721	20.818	.33	2.9097	2.8908	2.8721	2.8533	2.8350
.05	19.975	19.208	18.501	17.842	17.218	.34	2.8170	2.7992	2.7815	2.7641	2.7468
.06	16.645	16.111	15.598	15.127	14.682	.35	2.7295	2.7126	2.6960	2.6794	2.6629
.07	14.257	13.862	13.489	13.128	12.786	.36	2.6467	2.6308	2.6150	2.5991	2.5837
.08	12.472	12.163	11.874	11.598	11.330	.37	2.5684	2.5533	2.5380	2.5232	2.5086
.09	11.078	10.833	10.602	10.382	10.165	.38	2.4938	2.4794	2.4650	2.4509	2.4370
.10	9.9675	9.7664	9.5755	9.3944	9.2199	.39	2.4230	2.4093	2.3956	2.3821	2.3689
.11	9.0492	8.8870	8.7312	8.5777	8.4315	.40	2.3556	2.3426	2.3295	2.3167	2.3039
.12	8.2880	8.1513	8.0176	7.8894	7.7658	.41	2.2914	2.2788	2.2663	2.2541	2.2419
.13	7.6434	7.5272	7.4120	7.3024	7.1938	.42	2.2298	2.2180	2.2060	2.1944	2.1828
.14	7.0904	6.9898	6.8905	6.7953	6.7013	.43	2.1712	2.1599	2.1485	2.1372	2.1262
.15	6.6110	6.5218	6.4362	6.3514	6.2704	.44	2.1152	2.1042	2.0933	2.0827	2.0720
.16	6.1899	6.1123	6.0359	5.9621	5.8893	.45	2.0614	2.0509	2.0405	2.0303	2.0201
.17	5.8192	5.7493	5.6823	5.6158	5.5518	.46	2.0099	1.9999	1.9899	1.9800	1.9702
.18	5.4882	5.4274	5.3669	5.3072	5.2488	.47	1.9605	1.9508	1.9412	1.9317	1.9223
.19	5.1928	5.1367	5.0821	5.0291	4.9766	.48	1.9129	1.9037	1.8945	1.8854	1.8763
.20	4.9258	4.8751	4.8268	4.7777	4.7300	.49	1.8674	1.8585	1.8497	1.8409	1.8322
.21	4.6841	4.6382	4.5937	4.5494	4.5061	.50	1.8234	1.8149	1.8064	1.7980	1.7896
.22	4.4642	4.4222	4.3810	4.3413	4.3013	.51	1.7812	1.7729	1.7648	1.7567	1.7485
.23	4.2623	4.2247	4.1869	4.1497	4.1137	.52	1.7405	1.7325	1.7246	1.7167	1.7089
.24	4.0779	4.0425	4.0079	3.9743	3.9405	.53	1.7012	1.6936	1.6859	1.6783	1.6707
.25	3.9074	3.8748	3.8434	3.8117	3.7806	.54	1.6633	1.6559	1.6485	1.6412	1.6339
.26	3.7499	3.7203	3.6906	3.6612	3.6324	.55	1.6267	1.6195	1.6124	1.6053	1.5982
.27	3.6040	3.5764	3.5487	3.5216	3.4947	.56	1.5913	1.5843	1.5774	1.5707	1.5638
.28	3.4682	3.4422	3.4164	3.3915	3.3664	.57	1.5571	1.5503	1.5438	1.5372	1.5306

TABLE V.—Continued.

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.58	1.5240	1.5175	1.5110	1.5046	1.4982	1.29	.5215	.5208	.5202	.5195	.5189
.59	1.4919	1.4856	1.4794	1.4732	1.4670	1.30	.5182	.5176	.5170	.5164	.5158
.60	1.4609	1.4548	1.4488	1.4428	1.4368	1.31	.5153	.5147	.5141	.5136	.5131
.61	1.4309	1.4250	1.4191	1.4133	1.4075	1.32	.5125	.5120	.5115	.5110	.5105
.62	1.4017	1.3960	1.3903	1.3847	1.3791	1.33	.5101	.5096	.5091	.5087	.5083
.63	1.3734	1.3679	1.3624	1.3569	1.3515	1.34	.5079	.5075	.5070	.5067	.5063
.64	1.3460	1.3407	1.3353	1.3300	1.3247	1.35	.5059	.5056	.5052	.5049	.5045
.65	1.3193	1.3142	1.3089	1.3038	1.2986	1.36	.5042	.5039	.5036	.5034	.5031
.66	1.2935	1.2885	1.2834	1.2784	1.2733	1.37	.5028	.5026	.5023	.5021	.5019
.67	1.2684	1.2635	1.2586	1.2537	1.2488	1.38	.5017	.5015	.5013	.5011	.5010
.68	1.2440	1.2392	1.2344	1.2296	1.2249	1.39	.5008	.5007	.5006	.5005	.5004
.69	1.2202	1.2156	1.2109	1.2063	1.2017	1.40	.5003	.5002	.5001	.5001	.5000
.70	1.1972	1.1926	1.1881	1.1836	1.1791	1.41	.5000	.5000	.5000	.5000	.5000
.71	1.1747	1.1702	1.1659	1.1615	1.1571	1.42	.5000	.5001	.5001	.5002	.5003
.72	1.1528	1.1485	1.1442	1.1399	1.1357	1.43	.5004	.5005	.5006	.5007	.5008
.73	1.1315	1.1273	1.1232	1.1191	1.1149	1.44	.5010	.5012	.5013	.5015	.5017
.74	1.1108	1.1067	1.1027	1.0986	1.0946	1.45	.5019	.5022	.5024	.5027	.5030
.75	1.0907	1.0867	1.0827	1.0788	1.0748	1.46	.5032	.5035	.5038	.5042	.5045
.76	1.0709	1.0671	1.0632	1.0594	1.0556	1.47	.5048	.5052	.5056	.5060	.5064
.77	1.0518	1.0480	1.0443	1.0406	1.0369	1.48	.5068	.5072	.5077	.5081	.5086
.78	1.0331	1.0295	1.0258	1.0222	1.0185	1.49	.5091	.5096	.5101	.5107	.5112
.79	1.0150	1.0114	1.0078	1.0043	1.0007	1.50	.5118	.5124	.5130	.5136	.5142
.80	.9972	.9937	.9903	.9868	.9834	1.51	.5148	.5155	.5162	.5169	.5176
.81	.9800	.9765	.9731	.9698	.9664	1.52	.5183	.5190	.5198	.5205	.5213
.82	.9631	.9597	.9564	.9532	.9499	1.53	.5221	.5230	.5238	.5247	.5255
.83	.9467	.9434	.9402	.9370	.9338	1.54	.5264	.5273	.5282	.5292	.5301
.84	.9306	.9274	.9243	.9212	.9181	1.55	.5311	.5321	.5331	.5341	.5352
.85	.9149	.9119	.9088	.9057	.9027	1.56	.5362	.5373	.5384	.5396	.5407
.86	.8997	.8967	.8937	.8907	.8877	1.57	.5418	.5430	.5442	.5454	.5467
.87	.8848	.8819	.8789	.8761	.8732	1.58	.5479	.5492	.5505	.5518	.5532
.88	.8703	.8674	.8646	.8618	.8590	1.59	.5545	.5559	.5573	.5587	.5602
.89	.8561	.8534	.8506	.8478	.8450	1.60	.5616	.5631	.5647	.5662	.5677
.90	.8423	.8396	.8369	.8342	.8315	1.61	.5693	.5709	.5726	.5742	.5759
.91	.8288	.8262	.8235	.8209	.8183	1.62	.5776	.5793	.5811	.5828	.5846
.92	.8157	.8131	.8105	.8080	.8054	1.63	.5864	.5883	.5902	.5921	.5940
.93	.8029	.8003	.7978	.7953	.7929	1.64	.5960	.5980	.6000	.6020	.6041
.94	.7904	.7879	.7854	.7830	.7806	1.65	.6061	.6083	.6104	.6126	.6148
.95	.7782	.7758	.7734	.7710	.7687	1.66	.6171	.6194	.6217	.6240	.6264
.96	.7663	.7639	.7616	.7593	.7570	1.67	.6288	.6312	.6337	.6362	.6387
.97	.7547	.7524	.7501	.7479	.7456	1.68	.6412	.6438	.6465	.6491	.6518
.98	.7434	.7412	.7390	.7368	.7346	1.69	.6546	.6574	.6602	.6630	.6659
.99	.7324	.7302	.7281	.7259	.7238	1.70	.6688	.6718	.6749	.6779	.6810
1.00	.7217	.7195	.7175	.7153	.7133	1.71	.6841	.6873	.6905	.6938	.6971
1.01	.7112	.7091	.7071	.7051	.7031	1.72	.7005	.7039	.7073	.7108	.7143
1.02	.7010	.6991	.6971	.6951	.6931	1.73	.7179	.7216	.7253	.7290	.7328
1.03	.6912	.6892	.6873	.6853	.6834	1.74	.7367	.7405	.7445	.7485	.7526
1.04	.6815	.6796	.6777	.6758	.6740	1.75	.7567	.7609	.7652	.7694	.7738
1.05	.6722	.6703	.6685	.6667	.6648	1.76	.7782	.7827	.7873	.7919	.7966
1.06	.6630	.6612	.6595	.6577	.6559	1.77	.8014	.8062	.8111	.8161	.8212
1.07	.6542	.6524	.6507	.6490	.6473	1.78	.8263	.8315	.8368	.8422	.8476
1.08	.6456	.6439	.6422	.6406	.6389	1.79	.8531	.8588	.8645	.8703	.8762
1.09	.6372	.6356	.6340	.6324	.6308	1.80	.8822	.8883	.8945	.9008	.9072
1.10	.6292	.6276	.6260	.6249	.6229	1.81	.9137	.9203	.9270	.9338	.9409
1.11	.6213	.6198	.6182	.6167	.6152	1.82	.9479	.9551	.9624	.9699	.9775
1.12	.6137	.6122	.6107	.6092	.6078	1.83	.9852	.9931	1.0011	1.0093	1.0176
1.13	.6064	.6049	.6035	.6021	.6006	1.84	1.0261	1.0348	1.0436	1.0526	1.0618
1.14	.5993	.5978	.5965	.5951	.5937	1.85	1.0711	1.0807	1.0904	1.1003	1.1105
1.15	.5924	.5910	.5897	.5884	.5871	1.86	1.1209	1.1314	1.1423	1.1533	1.1647
1.16	.5857	.5844	.5832	.5819	.5806	1.87	1.1762	1.1881	1.2002	1.2126	1.2253
1.17	.5794	.5781	.5769	.5756	.5744	1.88	1.2384	1.2517	1.2654	1.2794	1.2939
1.18	.5732	.5720	.5708	.5696	.5685	1.89	1.3086	1.3238	1.3393	1.3555	1.3720
1.19	.5673	.5661	.5650	.5639	.5628	1.90	1.3889	1.4064	1.4244	1.4430	1.4621
1.20	.5616	.5605	.5594	.5583	.5573	1.91	1.4819	1.5023	1.5235	1.5452	1.5678
1.21	.5562	.5552	.5541	.5531	.5520	1.92	1.5913	1.6156	1.6408	1.6670	1.6942
1.22	.5510	.5500	.5490	.5480	.5470	1.93	1.7226	1.7521	1.7831	1.8153	1.8491
1.23	.5461	.5451	.5442	.5432	.5423	1.94	1.8845	1.9217	1.9608	2.0019	2.0451
1.24	.5414	.5405	.5396	.5387	.5378	1.95	2.0911	2.1399	2.1917	2.2469	2.3056
1.25	.5369	.5360	.5352	.5343	.5335	1.96	2.3689	2.4367	2.5103	2.5898	2.6767
1.26	.5327	.5319	.5311	.5303	.5295	1.97	2.7719	2.8766	2.9932	3.1239	3.2717
1.27	.5287	.5279	.5272	.5264	.5257	1.98	3.4403	3.6364	3.8671	4.1458	4.4898
1.28	.5250	.5243	.5235	.5229	.5222	1.99	4.9321	5.5296	6.4024	7.8633	11.152

TABLE VI. Polarization factor, $(p)^{\dagger} \left[= \left(\frac{1 + \cos^2 2\theta}{2} \right)^{\dagger} \right]$, as a function of $\sin \theta$.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.01	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.02	1.000	1.000	.999	.999	.999	.999	.999	.999	.999	.999
.03	.999	.999	.999	.999	.999	.999	.999	.999	.999	.998
.04	.998	.998	.998	.998	.998	.998	.998	.998	.998	.998
.05	.997	.997	.997	.997	.997	.997	.997	.997	.997	.997
.06	.996	.996	.996	.996	.996	.996	.996	.996	.995	.995
.07	.995	.995	.995	.995	.995	.994	.994	.994	.994	.994
.08	.994	.993	.993	.993	.993	.993	.993	.992	.992	.992
.09	.992	.992	.992	.991	.991	.991	.991	.991	.990	.990
.10	.990	.990	.990	.989	.989	.989	.989	.989	.988	.988
.11	.988	.988	.988	.987	.987	.987	.987	.986	.986	.986
.12	.986	.985	.985	.985	.985	.984	.984	.984	.984	.984
.13	.983	.983	.983	.982	.982	.982	.982	.981	.981	.981
.14	.981	.980	.980	.980	.979	.979	.979	.979	.978	.978
.15	.978	.977	.977	.977	.977	.976	.976	.976	.975	.975
.16	.975	.974	.974	.974	.973	.973	.973	.973	.972	.972
.17	.972	.971	.971	.971	.970	.970	.970	.969	.969	.969
.18	.968	.968	.967	.967	.967	.966	.966	.966	.965	.965
.19	.965	.964	.964	.963	.963	.963	.962	.962	.962	.961
.20	.961	.960	.960	.960	.959	.959	.958	.958	.958	.957
.21	.957	.957	.956	.956	.955	.955	.954	.954	.954	.953
.22	.953	.952	.952	.952	.951	.951	.950	.950	.949	.949
.23	.949	.948	.948	.947	.947	.946	.946	.946	.945	.945
.24	.944	.944	.943	.943	.942	.942	.941	.941	.941	.940
.25	.940	.939	.939	.938	.938	.937	.937	.936	.936	.935
.26	.935	.934	.934	.933	.933	.932	.932	.931	.931	.930
.27	.930	.929	.929	.928	.928	.927	.927	.926	.926	.925
.28	.925	.924	.924	.923	.923	.922	.922	.921	.921	.920
.29	.920	.919	.919	.918	.918	.917	.917	.916	.916	.915
.30	.914	.914	.913	.913	.912	.912	.911	.911	.910	.910
.31	.909	.908	.908	.907	.907	.906	.906	.905	.905	.904
.32	.903	.903	.902	.902	.901	.901	.900	.899	.899	.898
.33	.898	.897	.897	.896	.895	.895	.894	.894	.893	.893
.34	.892	.891	.891	.890	.890	.889	.888	.888	.887	.887
.35	.886	.885	.885	.884	.884	.883	.882	.882	.881	.881
.36	.880	.879	.879	.878	.878	.877	.876	.876	.875	.875
.37	.874	.873	.873	.872	.871	.871	.870	.870	.869	.868
.38	.868	.867	.866	.866	.865	.865	.864	.863	.863	.862
.39	.861	.861	.860	.860	.859	.858	.858	.857	.856	.856
.40	.855	.854	.854	.853	.853	.852	.851	.851	.850	.849
.41	.849	.848	.847	.847	.846	.846	.845	.844	.844	.843
.42	.842	.842	.841	.840	.840	.839	.838	.838	.837	.836
.43	.836	.835	.835	.834	.833	.833	.832	.831	.831	.830
.44	.829	.829	.828	.827	.827	.826	.825	.825	.824	.823
.45	.823	.822	.822	.821	.820	.820	.819	.818	.818	.817
.46	.816	.816	.815	.814	.814	.813	.812	.812	.811	.810
.47	.810	.809	.809	.808	.807	.807	.806	.805	.805	.804
.48	.803	.803	.802	.801	.801	.800	.800	.799	.798	.798
.49	.797	.796	.796	.795	.794	.794	.793	.792	.792	.791
.50	.791	.790	.789	.789	.788	.787	.787	.786	.785	.785
.51	.784	.784	.783	.782	.782	.781	.781	.780	.779	.779
.52	.778	.777	.777	.776	.776	.775	.774	.774	.773	.773
.53	.772	.771	.771	.770	.770	.769	.768	.768	.767	.767
.54	.766	.766	.765	.764	.764	.763	.763	.762	.761	.761
.55	.760	.760	.759	.759	.758	.757	.757	.756	.756	.755
.56	.755	.754	.754	.753	.752	.752	.751	.751	.750	.750
.57	.749	.749	.748	.748	.747	.747	.746	.746	.745	.745
.58	.744	.744	.743	.742	.742	.741	.741	.740	.740	.740
.59	.739	.739	.738	.738	.737	.737	.736	.736	.735	.735
.60	.734	.734	.733	.733	.732	.732	.732	.731	.731	.730
.61	.730	.729	.729	.729	.728	.728	.727	.727	.727	.726
.62	.726	.725	.725	.725	.724	.724	.723	.723	.723	.722
.63	.722	.722	.721	.721	.721	.720	.720	.720	.719	.719
.64	.719	.718	.718	.718	.717	.717	.717	.716	.716	.716
.65	.716	.715	.715	.715	.714	.714	.714	.714	.713	.713
.66	.713	.713	.712	.712	.712	.712	.712	.711	.711	.711
.67	.711	.710	.710	.710	.710	.710	.710	.710	.709	.709

TABLE VI.—Continued.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.68	.709	.709	.709	.709	.709	.708	.708	.708	.708	.708
.69	.708	.708	.708	.708	.708	.708	.707	.707	.707	.707
.70	.707	.707	.707	.707	.707	.707	.707	.707	.707	.707
.71	.707	.707	.707	.707	.707	.707	.707	.707	.707	.708
.72	.708	.708	.708	.708	.708	.708	.708	.708	.708	.709
.73	.709	.709	.709	.709	.709	.709	.710	.710	.710	.710
.74	.710	.710	.711	.711	.711	.711	.712	.712	.712	.712
.75	.713	.713	.713	.713	.714	.714	.714	.715	.715	.715
.76	.716	.716	.716	.717	.717	.717	.718	.718	.718	.719
.77	.719	.720	.720	.720	.721	.721	.722	.722	.723	.723
.78	.724	.724	.724	.725	.725	.726	.726	.727	.728	.728
.79	.729	.729	.730	.730	.731	.731	.732	.733	.733	.734
.80	.734	.735	.736	.736	.737	.737	.738	.739	.739	.740
.81	.741	.741	.742	.743	.744	.744	.745	.746	.746	.747
.82	.748	.749	.749	.750	.751	.752	.753	.753	.754	.755
.83	.756	.757	.758	.758	.759	.760	.761	.762	.763	.764
.84	.765	.765	.766	.767	.768	.769	.770	.771	.772	.773
.85	.774	.775	.776	.777	.778	.779	.780	.781	.782	.783
.86	.784	.785	.786	.787	.788	.789	.791	.792	.793	.794
.87	.795	.796	.797	.798	.800	.801	.802	.803	.804	.805
.88	.807	.808	.809	.810	.811	.813	.814	.815	.816	.818
.89	.819	.820	.821	.823	.824	.825	.827	.828	.829	.831
.90	.832	.833	.835	.836	.837	.839	.840	.842	.843	.844
.91	.846	.847	.849	.850	.851	.853	.854	.856	.857	.859
.92	.860	.862	.863	.865	.866	.868	.869	.871	.872	.874
.93	.875	.877	.879	.880	.882	.883	.885	.886	.888	.890
.94	.891	.893	.894	.896	.898	.899	.901	.903	.904	.906
.95	.908	.909	.911	.913	.915	.916	.918	.920	.921	.923
.96	.925	.927	.928	.930	.932	.934	.936	.937	.939	.941
.97	.943	.945	.946	.948	.950	.952	.954	.956	.957	.959
.98	.961	.963	.965	.967	.969	.971	.973	.975	.976	.978
.99	.980	.982	.984	.986	.988	.990	.992	.994	.996	.998

TABLE VII. Polarization factor, $(p)^{\frac{1}{2}} \left[= \left(\frac{1 + \cos^2 2\theta}{2} \right)^{\frac{1}{2}} \right]$, as a function of σ .

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.00	1.000	1.000	1.000	1.000	1.000	.31	.976	.976	.976	.975	.975
.01	1.000	1.000	1.000	1.000	1.000	.32	.975	.974	.974	.974	.973
.02	1.000	1.000	1.000	1.000	1.000	.33	.973	.973	.973	.972	.972
.03	1.000	1.000	1.000	1.000	1.000	.34	.972	.971	.971	.971	.970
.04	1.000	1.000	.999	.999	.999	.35	.970	.970	.969	.969	.969
.05	.999	.999	.999	.999	.999	.36	.968	.968	.967	.967	.967
.06	.999	.999	.999	.999	.999	.37	.966	.966	.966	.965	.965
.07	.999	.999	.999	.999	.998	.38	.965	.964	.964	.963	.963
.08	.998	.998	.998	.998	.998	.39	.963	.962	.962	.962	.961
.09	.998	.998	.998	.998	.998	.40	.961	.960	.960	.960	.959
.10	.997	.997	.997	.997	.997	.41	.959	.958	.958	.958	.957
.11	.997	.997	.997	.997	.997	.42	.957	.957	.956	.956	.955
.12	.996	.996	.996	.996	.996	.43	.955	.954	.954	.954	.953
.13	.996	.996	.996	.995	.995	.44	.953	.952	.952	.952	.951
.14	.995	.995	.995	.995	.995	.45	.951	.950	.950	.949	.949
.15	.994	.994	.994	.994	.994	.46	.949	.948	.948	.947	.947
.16	.994	.993	.993	.993	.993	.47	.946	.946	.946	.945	.945
.17	.993	.993	.992	.992	.992	.48	.944	.944	.943	.943	.942
.18	.992	.992	.992	.991	.991	.49	.942	.941	.941	.941	.940
.19	.991	.991	.991	.990	.990	.50	.940	.939	.939	.938	.938
.20	.990	.990	.990	.989	.989	.51	.937	.937	.936	.936	.935
.21	.989	.989	.989	.988	.988	.52	.935	.934	.934	.933	.933
.22	.988	.988	.988	.987	.987	.53	.932	.932	.931	.931	.930
.23	.987	.987	.986	.986	.986	.54	.930	.929	.929	.928	.928
.24	.986	.985	.985	.985	.985	.55	.927	.927	.926	.926	.925
.25	.984	.984	.984	.984	.984	.56	.925	.924	.924	.923	.923
.26	.983	.983	.983	.982	.982	.57	.922	.922	.921	.921	.920
.27	.982	.982	.981	.981	.981	.58	.920	.919	.919	.918	.918
.28	.981	.980	.980	.980	.979	.59	.917	.917	.916	.916	.915
.29	.979	.979	.979	.978	.978	.60	.914	.914	.913	.913	.912
.30	.978	.977	.977	.977	.977	.61	.912	.911	.911	.910	.910

TABLE VII.—Continued.

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.62	.909	.908	.908	.907	.907	1.31	.714	.714	.714	.713	.713
.63	.906	.906	.905	.905	.904	1.32	.713	.713	.712	.712	.712
.64	.903	.903	.902	.902	.901	1.33	.712	.712	.711	.711	.711
.65	.901	.900	.899	.899	.898	1.34	.711	.710	.710	.710	.710
.66	.898	.897	.897	.896	.895	1.35	.710	.710	.710	.709	.709
.67	.895	.894	.894	.893	.893	1.36	.709	.709	.709	.709	.709
.68	.892	.891	.891	.890	.890	1.37	.708	.708	.708	.708	.708
.69	.889	.888	.888	.887	.887	1.38	.708	.708	.708	.708	.708
.70	.886	.885	.885	.884	.884	1.39	.708	.707	.707	.707	.707
.71	.883	.882	.882	.881	.881	1.40	.707	.707	.707	.707	.707
.72	.880	.879	.879	.878	.878	1.41	.707	.707	.707	.707	.707
.73	.877	.876	.876	.875	.875	1.42	.707	.707	.707	.707	.707
.74	.874	.873	.873	.872	.871	1.43	.707	.707	.707	.707	.708
.75	.871	.870	.870	.869	.868	1.44	.708	.708	.708	.708	.708
.76	.868	.867	.866	.866	.865	1.45	.708	.708	.708	.708	.709
.77	.865	.864	.863	.863	.862	1.46	.709	.709	.709	.709	.709
.78	.861	.861	.860	.860	.859	1.47	.709	.710	.710	.710	.710
.79	.858	.858	.857	.856	.856	1.48	.710	.710	.711	.711	.711
.80	.855	.854	.854	.853	.853	1.49	.711	.712	.712	.712	.712
.81	.852	.851	.851	.850	.849	1.50	.713	.713	.713	.713	.714
.82	.849	.848	.847	.847	.846	1.51	.714	.714	.715	.715	.715
.83	.846	.845	.844	.844	.843	1.52	.716	.716	.716	.717	.717
.84	.842	.842	.841	.840	.840	1.53	.717	.718	.718	.718	.719
.85	.839	.838	.838	.837	.836	1.54	.719	.720	.720	.720	.721
.86	.836	.835	.835	.834	.833	1.55	.721	.722	.722	.723	.723
.87	.833	.832	.831	.831	.830	1.56	.724	.724	.724	.725	.725
.88	.829	.829	.828	.827	.827	1.57	.726	.726	.727	.728	.728
.89	.826	.825	.825	.824	.823	1.58	.729	.729	.730	.730	.731
.90	.823	.822	.822	.821	.820	1.59	.731	.732	.733	.733	.734
.91	.820	.819	.818	.818	.817	1.60	.734	.735	.736	.736	.737
.92	.816	.816	.815	.814	.814	1.61	.737	.738	.739	.739	.740
.93	.813	.812	.812	.811	.810	1.62	.741	.741	.742	.743	.744
.94	.810	.809	.809	.808	.807	1.63	.744	.745	.746	.746	.747
.95	.807	.806	.805	.805	.804	1.64	.748	.749	.749	.750	.751
.96	.803	.803	.802	.801	.801	1.65	.752	.753	.753	.754	.755
.97	.800	.800	.799	.798	.798	1.66	.756	.757	.758	.758	.759
.98	.797	.796	.796	.795	.794	1.67	.760	.761	.762	.763	.764
.99	.794	.793	.792	.792	.791	1.68	.765	.765	.766	.767	.768
1.00	.791	.790	.789	.789	.788	1.69	.769	.770	.771	.772	.773
1.01	.787	.787	.786	.785	.785	1.70	.774	.775	.776	.777	.778
1.02	.784	.784	.783	.782	.782	1.71	.779	.780	.781	.782	.783
1.03	.781	.781	.780	.779	.779	1.72	.784	.785	.786	.787	.788
1.04	.778	.777	.777	.776	.776	1.73	.789	.791	.792	.793	.794
1.05	.775	.774	.774	.773	.773	1.74	.795	.796	.797	.798	.800
1.06	.772	.771	.771	.770	.770	1.75	.801	.802	.803	.804	.805
1.07	.769	.768	.768	.767	.767	1.76	.807	.808	.809	.810	.811
1.08	.766	.766	.765	.764	.764	1.77	.813	.814	.815	.816	.818
1.09	.763	.763	.762	.761	.761	1.78	.819	.820	.821	.823	.824
1.10	.760	.760	.759	.759	.758	1.79	.825	.827	.828	.829	.831
1.11	.757	.757	.756	.756	.755	1.80	.832	.833	.835	.836	.837
1.12	.755	.754	.754	.753	.752	1.81	.839	.840	.842	.843	.844
1.13	.752	.751	.751	.750	.750	1.82	.846	.847	.849	.850	.851
1.14	.749	.749	.748	.748	.747	1.83	.853	.854	.856	.857	.859
1.15	.747	.746	.746	.745	.745	1.84	.860	.862	.863	.865	.866
1.16	.744	.744	.743	.742	.742	1.85	.868	.869	.871	.872	.874
1.17	.741	.741	.740	.740	.740	1.86	.875	.877	.879	.880	.882
1.18	.739	.739	.738	.737	.737	1.87	.883	.885	.886	.888	.890
1.19	.737	.736	.736	.735	.735	1.88	.891	.893	.894	.896	.898
1.20	.734	.734	.733	.733	.732	1.89	.899	.901	.903	.904	.906
1.21	.732	.732	.731	.731	.730	1.90	.908	.909	.911	.913	.915
1.22	.730	.729	.729	.729	.728	1.91	.916	.918	.920	.921	.923
1.23	.728	.727	.727	.727	.726	1.92	.925	.927	.928	.930	.932
1.24	.726	.725	.725	.725	.724	1.93	.934	.936	.937	.939	.941
1.25	.724	.723	.723	.723	.722	1.94	.943	.945	.946	.948	.950
1.26	.722	.722	.721	.721	.721	1.95	.952	.954	.956	.957	.959
1.27	.720	.720	.720	.719	.719	1.96	.961	.963	.965	.967	.969
1.28	.719	.718	.718	.718	.717	1.97	.971	.973	.975	.976	.978
1.29	.717	.717	.716	.716	.716	1.98	.980	.982	.984	.986	.988
1.30	.716	.715	.715	.715	.714	1.99	.990	.992	.994	.996	.998

TABLE VIII. Unscaled part of Lorentz factor, $(S \cdot L)^{\dagger} [= (\csc T)^{\dagger}]$, as a function of T .

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	∞	23.09	16.09	13.08	11.10	10.07	9.772	9.047	8.463	7.979
1	7.570	7.217	6.910	6.639	6.398	6.181	5.985	5.806	5.643	5.492
2	5.353	5.224	5.123	4.992	4.887	4.789	4.695	4.608	4.524	4.446
3	4.371	4.300	4.232	4.168	4.106	4.047	3.991	3.937	3.885	3.834
4	3.787	3.740	3.695	3.652	3.611	3.571	3.531	3.493	3.457	3.422
5	3.387	3.354	3.321	3.291	3.260	3.230	3.202	3.173	3.146	3.119
6	3.093	3.068	3.043	3.019	2.995	2.972	2.950	2.928	2.906	2.885
7	2.865	2.844	2.825	2.805	2.786	2.768	2.750	2.732	2.714	2.697
8	2.680	2.664	2.648	2.632	2.616	2.601	2.586	2.571	2.557	2.542
9	2.523	2.515	2.501	2.488	2.474	2.462	2.449	2.436	2.424	2.412
10	2.400	2.388	2.376	2.365	2.354	2.342	2.332	2.321	2.310	2.300
11	2.289	2.279	2.269	2.259	2.249	2.240	2.230	2.221	2.211	2.202
12	2.193	2.184	2.175	2.167	2.158	2.149	2.141	2.133	2.125	2.116
13	2.108	2.100	2.093	2.085	2.077	2.070	2.062	2.055	2.047	2.040
14	2.033	2.026	2.019	2.012	2.005	1.998	1.992	1.985	1.979	1.972
15	1.966	1.959	1.953	1.947	1.941	1.934	1.928	1.922	1.917	1.910
16	1.905	1.899	1.893	1.888	1.882	1.876	1.871	1.865	1.860	1.855
17	1.849	1.844	1.839	1.834	1.829	1.824	1.819	1.814	1.809	1.804
18	1.799	1.794	1.789	1.785	1.780	1.775	1.771	1.766	1.762	1.757
19	1.753	1.748	1.744	1.740	1.735	1.731	1.727	1.722	1.718	1.714
20	1.710	1.706	1.702	1.698	1.694	1.690	1.686	1.682	1.678	1.674
21	1.670	1.667	1.663	1.659	1.656	1.652	1.648	1.645	1.641	1.637
22	1.634	1.630	1.627	1.623	1.620	1.616	1.613	1.610	1.607	1.603
23	1.600	1.597	1.593	1.590	1.587	1.584	1.581	1.577	1.574	1.571
24	1.568	1.565	1.562	1.559	1.556	1.553	1.550	1.547	1.544	1.541
25	1.538	1.535	1.533	1.530	1.527	1.524	1.521	1.519	1.516	1.513
26	1.510	1.508	1.505	1.502	1.500	1.497	1.494	1.492	1.489	1.487
27	1.484	1.482	1.479	1.476	1.474	1.472	1.469	1.467	1.464	1.462
28	1.459	1.457	1.455	1.452	1.450	1.448	1.445	1.443	1.441	1.438
29	1.436	1.434	1.432	1.429	1.427	1.425	1.423	1.421	1.418	1.416
30	1.414	1.412	1.410	1.408	1.406	1.404	1.402	1.400	1.397	1.395
31	1.393	1.391	1.389	1.387	1.385	1.383	1.381	1.379	1.377	1.375
32	1.374	1.372	1.370	1.368	1.366	1.364	1.362	1.361	1.359	1.357
33	1.355	1.353	1.351	1.349	1.348	1.346	1.344	1.342	1.341	1.339
34	1.337	1.336	1.334	1.332	1.330	1.329	1.327	1.326	1.324	1.322
35	1.320	1.319	1.317	1.316	1.314	1.312	1.311	1.309	1.308	1.306
36	1.304	1.303	1.301	1.300	1.298	1.297	1.295	1.293	1.292	1.291
37	1.289	1.288	1.286	1.285	1.283	1.282	1.280	1.279	1.277	1.276
38	1.274	1.273	1.272	1.270	1.269	1.267	1.266	1.265	1.263	1.262
39	1.261	1.259	1.258	1.257	1.255	1.254	1.253	1.251	1.250	1.249
40	1.247	1.246	1.245	1.243	1.242	1.241	1.240	1.239	1.237	1.236
41	1.235	1.233	1.232	1.231	1.230	1.228	1.227	1.226	1.225	1.224
42	1.222	1.221	1.220	1.219	1.218	1.217	1.215	1.214	1.213	1.212
43	1.211	1.210	1.209	1.207	1.206	1.205	1.204	1.203	1.202	1.201
44	1.200	1.199	1.197	1.196	1.195	1.194	1.193	1.192	1.191	1.190
45	1.189	1.188	1.187	1.186	1.185	1.184	1.183	1.182	1.181	1.180
46	1.179	1.178	1.177	1.176	1.175	1.174	1.173	1.172	1.171	1.170
47	1.169	1.168	1.167	1.167	1.166	1.165	1.164	1.163	1.162	1.161
48	1.160	1.159	1.158	1.157	1.156	1.155	1.155	1.154	1.153	1.152
49	1.151	1.150	1.149	1.148	1.148	1.147	1.146	1.145	1.144	1.143
50	1.142	1.142	1.141	1.140	1.139	1.138	1.138	1.137	1.136	1.135
51	1.134	1.134	1.133	1.132	1.131	1.130	1.130	1.129	1.128	1.127
52	1.126	1.126	1.125	1.124	1.123	1.122	1.122	1.121	1.120	1.120
53	1.119	1.118	1.118	1.117	1.116	1.115	1.114	1.114	1.113	1.113
54	1.112	1.111	1.110	1.110	1.109	1.108	1.108	1.107	1.106	1.105
55	1.105	1.104	1.104	1.103	1.102	1.101	1.101	1.100	1.100	1.099
56	1.098	1.098	1.097	1.096	1.096	1.095	1.095	1.094	1.093	1.093
57	1.092	1.091	1.091	1.090	1.089	1.089	1.088	1.088	1.087	1.086
58	1.086	1.085	1.085	1.084	1.084	1.083	1.083	1.082	1.081	1.081
59	1.080	1.079	1.079	1.078	1.078	1.077	1.077	1.076	1.076	1.075
60	1.075	1.074	1.073	1.073	1.072	1.072	1.071	1.071	1.070	1.070
61	1.069	1.069	1.068	1.068	1.067	1.067	1.066	1.066	1.065	1.065
62	1.064	1.064	1.063	1.063	1.062	1.062	1.061	1.061	1.060	1.060
63	1.059	1.059	1.058	1.058	1.057	1.057	1.056	1.056	1.056	1.055
64	1.055	1.055	1.054	1.054	1.053	1.053	1.052	1.052	1.051	1.051
65	1.050	1.050	1.050	1.049	1.049	1.048	1.048	1.047	1.047	1.046
66	1.046	1.046	1.045	1.045	1.045	1.044	1.044	1.044	1.043	1.043
67	1.042	1.042	1.042	1.041	1.041	1.040	1.040	1.040	1.039	1.039
68	1.039	1.038	1.038	1.037	1.037	1.037	1.036	1.036	1.036	1.035

TABLE VIII.—Continued.

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
69	1.035	1.034	1.034	1.034	1.033	1.033	1.033	1.032	1.032	1.032
70	1.031	1.031	1.031	1.031	1.031	1.030	1.030	1.030	1.029	1.029
71	1.029	1.028	1.028	1.028	1.027	1.027	1.027	1.026	1.026	1.026
72	1.025	1.025	1.025	1.025	1.024	1.024	1.024	1.023	1.023	1.023
73	1.023	1.022	1.022	1.022	1.021	1.021	1.021	1.021	1.020	1.020
74	1.020	1.020	1.019	1.019	1.019	1.019	1.018	1.018	1.018	1.018
75	1.017	1.017	1.017	1.017	1.016	1.016	1.016	1.016	1.016	1.015
76	1.015	1.015	1.015	1.014	1.014	1.014	1.014	1.014	1.013	1.013
77	1.013	1.013	1.012	1.012	1.012	1.012	1.012	1.011	1.011	1.011
78	1.011	1.011	1.011	1.010	1.010	1.010	1.010	1.010	1.009	1.009
79	1.009	1.009	1.009	1.009	1.008	1.008	1.008	1.008	1.008	1.008
80	1.007	1.007	1.007	1.007	1.007	1.007	1.007	1.006	1.006	1.006
81	1.006	1.006	1.006	1.006	1.005	1.005	1.005	1.005	1.005	1.005
82	1.005	1.005	1.004	1.004	1.004	1.004	1.004	1.004	1.004	1.004
83	1.004	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.003
84	1.003	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002
85	1.002	1.002	1.002	1.001	1.001	1.001	1.001	1.001	1.001	1.001
86	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.000
87	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
88	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
89	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
91	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
92	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
93	1.000	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
94	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.002	1.002	1.002
95	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.003
96	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.004
97	1.004	1.004	1.004	1.004	1.004	1.004	1.004	1.004	1.005	1.005
98	1.005	1.005	1.005	1.005	1.005	1.005	1.006	1.006	1.006	1.006
99	1.006	1.006	1.006	1.007	1.007	1.007	1.007	1.007	1.007	1.007
100	1.008	1.008	1.008	1.008	1.008	1.008	1.009	1.009	1.009	1.009
101	1.009	1.009	1.010	1.010	1.010	1.010	1.010	1.011	1.011	1.011
102	1.011	1.011	1.011	1.012	1.012	1.012	1.012	1.012	1.013	1.013
103	1.013	1.013	1.014	1.014	1.014	1.014	1.014	1.015	1.015	1.015
104	1.015	1.016	1.016	1.016	1.016	1.016	1.017	1.017	1.017	1.017
105	1.018	1.018	1.018	1.018	1.019	1.019	1.019	1.019	1.020	1.020
106	1.020	1.020	1.021	1.021	1.021	1.021	1.022	1.022	1.022	1.023
107	1.023	1.023	1.023	1.024	1.024	1.024	1.025	1.025	1.025	1.025
108	1.026	1.026	1.026	1.027	1.027	1.027	1.028	1.028	1.028	1.029
109	1.029	1.029	1.030	1.030	1.030	1.031	1.031	1.031	1.031	1.031
110	1.032	1.032	1.032	1.033	1.033	1.033	1.034	1.034	1.034	1.035
111	1.035	1.036	1.036	1.036	1.037	1.037	1.037	1.038	1.038	1.039
112	1.039	1.039	1.040	1.040	1.040	1.041	1.041	1.042	1.042	1.042
113	1.043	1.043	1.044	1.044	1.044	1.045	1.045	1.045	1.046	1.046
114	1.046	1.047	1.047	1.048	1.048	1.049	1.049	1.050	1.050	1.050
115	1.051	1.051	1.052	1.052	1.053	1.053	1.054	1.054	1.055	1.055
116	1.055	1.056	1.056	1.056	1.057	1.057	1.058	1.058	1.059	1.059
117	1.060	1.060	1.061	1.061	1.062	1.062	1.063	1.063	1.064	1.064
118	1.065	1.065	1.066	1.066	1.067	1.067	1.068	1.068	1.069	1.069
119	1.070	1.070	1.071	1.071	1.072	1.072	1.073	1.073	1.074	1.075
120	1.075	1.076	1.076	1.077	1.077	1.078	1.078	1.079	1.079	1.080
121	1.081	1.081	1.082	1.083	1.083	1.084	1.084	1.085	1.085	1.086
122	1.086	1.087	1.088	1.088	1.089	1.089	1.090	1.091	1.091	1.092
123	1.093	1.093	1.094	1.095	1.095	1.096	1.096	1.097	1.098	1.098
124	1.099	1.100	1.100	1.101	1.101	1.102	1.103	1.104	1.104	1.105
125	1.105	1.106	1.107	1.108	1.108	1.109	1.110	1.110	1.111	1.112
126	1.113	1.113	1.114	1.114	1.115	1.116	1.117	1.118	1.118	1.119
127	1.120	1.120	1.121	1.122	1.122	1.123	1.124	1.125	1.126	1.126
128	1.127	1.128	1.129	1.130	1.130	1.131	1.132	1.133	1.134	1.134
129	1.135	1.136	1.137	1.138	1.138	1.139	1.140	1.141	1.142	1.142
130	1.143	1.144	1.145	1.146	1.147	1.148	1.148	1.149	1.150	1.151
131	1.152	1.153	1.154	1.155	1.155	1.156	1.157	1.158	1.159	1.160
132	1.161	1.162	1.163	1.164	1.165	1.166	1.167	1.167	1.168	1.169
133	1.170	1.171	1.172	1.173	1.174	1.175	1.176	1.177	1.178	1.179
134	1.180	1.181	1.182	1.183	1.184	1.185	1.186	1.187	1.188	1.189
135	1.190	1.191	1.192	1.193	1.194	1.195	1.196	1.197	1.199	1.200
136	1.201	1.202	1.203	1.204	1.205	1.206	1.207	1.209	1.210	1.211
137	1.212	1.213	1.214	1.215	1.217	1.218	1.219	1.220	1.221	1.222

TABLE VIII.—Continued.

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
138	1.224	1.225	1.226	1.227	1.228	1.230	1.231	1.232	1.233	1.235
139	1.236	1.237	1.239	1.240	1.241	1.242	1.243	1.245	1.246	1.247
140	1.249	1.250	1.251	1.253	1.254	1.255	1.257	1.258	1.259	1.261
141	1.262	1.263	1.265	1.266	1.267	1.269	1.270	1.272	1.273	1.274
142	1.276	1.277	1.279	1.280	1.282	1.283	1.285	1.286	1.288	1.289
143	1.291	1.292	1.293	1.295	1.297	1.298	1.300	1.301	1.303	1.304
144	1.306	1.308	1.309	1.311	1.312	1.314	1.316	1.317	1.319	1.320
145	1.322	1.324	1.326	1.327	1.329	1.330	1.332	1.334	1.336	1.337
146	1.339	1.341	1.342	1.344	1.346	1.348	1.349	1.351	1.353	1.355
147	1.357	1.359	1.361	1.362	1.364	1.366	1.368	1.370	1.372	1.374
148	1.375	1.377	1.379	1.381	1.383	1.385	1.387	1.389	1.391	1.393
149	1.395	1.397	1.400	1.402	1.404	1.406	1.408	1.410	1.412	1.414
150	1.416	1.418	1.421	1.423	1.425	1.427	1.429	1.432	1.434	1.436
151	1.438	1.441	1.443	1.445	1.448	1.450	1.452	1.455	1.457	1.459
152	1.462	1.464	1.467	1.469	1.472	1.474	1.476	1.479	1.482	1.484
153	1.487	1.489	1.492	1.494	1.497	1.500	1.502	1.505	1.508	1.510
154	1.513	1.516	1.519	1.521	1.524	1.527	1.530	1.533	1.535	1.538
155	1.541	1.544	1.547	1.550	1.553	1.556	1.559	1.562	1.565	1.568
156	1.571	1.574	1.577	1.581	1.584	1.587	1.590	1.593	1.597	1.600
157	1.603	1.607	1.610	1.613	1.616	1.620	1.623	1.627	1.630	1.634
158	1.637	1.641	1.645	1.648	1.652	1.656	1.659	1.663	1.667	1.670
159	1.674	1.678	1.682	1.686	1.690	1.694	1.698	1.702	1.706	1.710
160	1.714	1.718	1.722	1.727	1.731	1.735	1.740	1.744	1.748	1.753
161	1.757	1.762	1.766	1.771	1.775	1.780	1.785	1.789	1.794	1.799
162	1.804	1.809	1.814	1.819	1.824	1.829	1.834	1.839	1.844	1.849
163	1.855	1.860	1.865	1.871	1.876	1.882	1.888	1.893	1.899	1.905
164	1.910	1.917	1.922	1.928	1.934	1.941	1.947	1.953	1.959	1.966
165	1.972	1.979	1.985	1.992	1.998	2.005	2.012	2.019	2.026	2.033
166	2.040	2.047	2.055	2.062	2.070	2.077	2.085	2.093	2.100	2.108
167	2.116	2.125	2.133	2.141	2.149	2.158	2.167	2.175	2.184	2.193
168	2.202	2.211	2.221	2.230	2.240	2.249	2.259	2.269	2.279	2.289
169	2.300	2.310	2.321	2.332	2.342	2.354	2.365	2.376	2.388	2.400
170	2.412	2.424	2.436	2.449	2.462	2.474	2.488	2.501	2.515	2.528
171	2.542	2.557	2.571	2.586	2.601	2.616	2.632	2.648	2.664	2.680
172	2.697	2.714	2.732	2.750	2.768	2.786	2.805	2.825	2.844	2.865
173	2.885	2.906	2.928	2.950	2.972	2.995	3.019	3.043	3.068	3.093
174	3.119	3.146	3.173	3.202	3.230	3.260	3.291	3.321	3.354	3.387
175	3.422	3.457	3.493	3.531	3.571	3.611	3.652	3.695	3.740	3.787
176	3.834	3.885	3.937	3.991	4.047	4.106	4.168	4.232	4.300	4.371
177	4.446	4.524	4.608	4.695	4.789	4.887	4.992	5.123	5.224	5.353
178	5.492	5.643	5.806	5.985	6.181	6.398	6.639	6.910	7.217	7.570
179	7.979	8.463	9.047	9.772	10.07	11.10	13.08	16.09	23.09	∞

the computation reduces to the following comparatively simple form

$$I \sim \{C_I \bar{A}_I + \{C_{II} \bar{A}_{II} + C_{III} \bar{A}_{III} + \dots\}^2 + \{C_I \bar{B}_I + \{C_{II} \bar{B}_{II} + C_{III} \bar{B}_{III}\}^2\}^2 \quad (15)$$

In the fortunate case that the crystal has inversion centers and that the origin is chosen at one of these, the right-hand column of (15), which represents the imaginary component, vanishes, and the computation can be further simplified to

$$(I)^{\dagger} \sim C_I \bar{A}_I + C_2 \bar{A}_2 + C_3 \bar{A}_3$$

(16)

The above computing scheme is discussed in greater detail elsewhere.⁴

APPENDIX

In an earlier paper we presented tables which simplified the computation of σ , which is the argument required when looking up the polarization factor of a reflection. We take this opportunity to present two other tables which simplify this computation further for the particular case of zero level reflections of tetragonal and hexagonal crystals. For these cases,

$$\text{Tetragonal: } \sigma_0 = \xi = a^*(h^2 + k^2)^{\dagger}, \quad (17)$$

$$\text{Hexagonal: } \sigma_0 = \xi = a^*(h^2 + k^2 + hk)^{\dagger}. \quad (18)$$

The functions $(h^2 + k^2)^{\dagger}$ and $(h^2 + k^2 + hk)^{\dagger}$ are given in Tables XI and XII

TABLE IX. Lorentz-polarization factor, $(L \cdot p)^2 \left[= \left(\frac{1}{\sin 2\theta} \cdot \frac{1 + \cos^2 2\theta}{2} \right)^2 \right]$, as a function of $\sin \theta$.

Valid only for zero level when $\mu = 0$.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.00	∞	22.42	15.79	12.91	11.21	10.01	9.129	8.442	7.908	7.451
.01	7.070	6.742	6.453	6.201	5.976	5.774	5.588	5.422	5.271	5.128
.02	5.000	4.879	4.766	4.660	4.563	4.470	4.383	4.301	4.224	4.150
.03	4.080	4.014	3.950	3.890	3.831	3.776	3.723	3.673	3.624	3.576
.04	3.531	3.487	3.445	3.406	3.366	3.329	3.291	3.256	3.222	3.189
.05	3.157	3.125	3.095	3.065	3.036	3.008	2.981	2.955	2.939	2.904
.06	2.889	2.855	2.832	2.809	2.787	2.765	2.744	2.722	2.702	2.682
.07	2.663	2.644	2.623	2.607	2.589	2.571	2.554	2.537	2.520	2.504
.08	2.488	2.472	2.457	2.442	2.427	2.412	2.398	2.384	2.370	2.356
.09	2.343	2.330	2.317	2.304	2.291	2.279	2.266	2.254	2.243	2.231
.10	2.219	2.208	2.197	2.186	2.175	2.164	2.154	2.143	2.133	2.123
.11	2.113	2.103	2.093	2.084	2.074	2.064	2.055	2.046	2.037	2.028
.12	2.019	2.011	2.002	1.993	1.985	1.977	1.969	1.960	1.952	1.944
.13	1.936	1.929	1.921	1.913	1.906	1.898	1.891	1.884	1.877	1.869
.14	1.862	1.855	1.848	1.842	1.835	1.828	1.821	1.815	1.808	1.802
.15	1.795	1.789	1.783	1.777	1.770	1.764	1.758	1.752	1.746	1.740
.16	1.734	1.729	1.723	1.717	1.711	1.706	1.700	1.695	1.689	1.684
.17	1.678	1.673	1.668	1.663	1.657	1.652	1.647	1.642	1.637	1.632
.18	1.627	1.622	1.617	1.612	1.607	1.602	1.598	1.593	1.588	1.584
.19	1.579	1.574	1.570	1.566	1.561	1.557	1.552	1.548	1.543	1.539
.20	1.535	1.531	1.526	1.522	1.518	1.514	1.510	1.505	1.501	1.497
.21	1.493	1.489	1.485	1.481	1.477	1.473	1.470	1.466	1.462	1.458
.22	1.454	1.451	1.447	1.443	1.439	1.436	1.432	1.429	1.425	1.421
.23	1.418	1.414	1.411	1.407	1.404	1.400	1.397	1.393	1.390	1.386
.24	1.383	1.380	1.377	1.373	1.370	1.366	1.363	1.360	1.357	1.354
.25	1.350	1.347	1.344	1.341	1.338	1.335	1.332	1.329	1.326	1.322
.26	1.319	1.316	1.313	1.310	1.307	1.304	1.301	1.298	1.295	1.293
.27	1.290	1.287	1.284	1.281	1.278	1.276	1.273	1.270	1.267	1.264
.28	1.261	1.259	1.256	1.253	1.251	1.248	1.245	1.243	1.240	1.237
.29	1.235	1.232	1.229	1.227	1.224	1.221	1.219	1.216	1.214	1.211
.30	1.209	1.206	1.204	1.201	1.199	1.196	1.194	1.191	1.189	1.187
.31	1.184	1.182	1.179	1.177	1.174	1.172	1.170	1.167	1.165	1.163
.32	1.160	1.158	1.155	1.153	1.151	1.148	1.146	1.144	1.142	1.140
.33	1.138	1.135	1.133	1.130	1.128	1.126	1.124	1.122	1.120	1.118
.34	1.115	1.113	1.111	1.109	1.107	1.105	1.103	1.100	1.098	1.096
.35	1.094	1.092	1.090	1.088	1.086	1.084	1.082	1.080	1.078	1.076
.36	1.074	1.072	1.070	1.068	1.066	1.064	1.062	1.060	1.058	1.056
.37	1.054	1.052	1.050	1.048	1.046	1.045	1.043	1.041	1.039	1.037
.38	1.035	1.033	1.031	1.029	1.028	1.026	1.024	1.022	1.020	1.018
.39	1.016	1.015	1.013	1.011	1.009	1.007	1.005	1.004	1.002	1.000
.40	.999	.997	.995	.993	.992	.990	.988	.986	.985	.983
.41	.981	.980	.978	.976	.975	.973	.971	.970	.968	.966
.42	.965	.963	.961	.960	.958	.957	.955	.953	.952	.950
.43	.949	.947	.945	.944	.942	.941	.939	.937	.936	.934
.44	.933	.931	.930	.928	.927	.925	.924	.922	.921	.919
.45	.918	.916	.915	.913	.912	.910	.909	.907	.906	.905
.46	.903	.902	.900	.899	.897	.896	.895	.893	.892	.890
.47	.889	.888	.886	.885	.884	.882	.881	.879	.878	.877
.48	.875	.874	.873	.871	.870	.869	.867	.866	.865	.863
.49	.862	.861	.860	.858	.857	.856	.855	.853	.852	.851
.50	.850	.848	.847	.846	.845	.843	.842	.841	.840	.839
.51	.837	.836	.835	.834	.833	.831	.830	.829	.828	.827
.52	.826	.824	.823	.822	.821	.820	.819	.818	.817	.815
.53	.814	.813	.812	.811	.810	.809	.808	.807	.806	.805
.54	.803	.802	.801	.800	.799	.798	.797	.796	.795	.794
.55	.793	.792	.791	.790	.789	.788	.787	.786	.785	.784
.56	.783	.782	.781	.781	.780	.779	.778	.777	.776	.775
.57	.774	.773	.772	.771	.771	.770	.769	.768	.767	.766
.58	.765	.764	.764	.763	.762	.761	.760	.760	.759	.758
.59	.757	.756	.756	.755	.754	.753	.752	.752	.751	.750
.60	.749	.749	.748	.747	.747	.746	.745	.744	.744	.743
.61	.742	.742	.741	.740	.740	.739	.738	.738	.737	.736
.62	.736	.735	.735	.734	.733	.733	.732	.732	.731	.730
.63	.730	.729	.729	.728	.728	.727	.727	.726	.726	.725
.64	.725	.724	.724	.723	.723	.722	.722	.721	.721	.720
.65	.720	.719	.719	.719	.718	.718	.717	.717	.717	.716
.66	.716	.716	.715	.715	.714	.714	.714	.714	.713	.713

TABLE IX.—Continued.

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.67	.713	.712	.712	.712	.712	.711	.711	.711	.711	.710
.68	.710	.710	.710	.710	.709	.709	.709	.709	.709	.708
.69	.708	.708	.708	.708	.708	.708	.708	.708	.707	.707
.70	.707	.707	.707	.707	.707	.707	.707	.707	.707	.707
.71	.707	.707	.707	.707	.707	.707	.707	.708	.708	.708
.72	.708	.708	.708	.708	.708	.708	.709	.709	.709	.709
.73	.709	.710	.710	.710	.710	.710	.711	.711	.711	.712
.74	.712	.712	.713	.713	.713	.714	.714	.714	.715	.715
.75	.715	.716	.716	.717	.717	.717	.718	.718	.719	.719
.76	.720	.720	.721	.721	.722	.723	.723	.724	.724	.725
.77	.726	.726	.727	.727	.728	.729	.729	.730	.731	.732
.78	.732	.733	.734	.735	.735	.736	.737	.738	.739	.739
.79	.740	.741	.742	.743	.744	.745	.746	.747	.747	.748
.80	.749	.750	.751	.752	.753	.755	.756	.757	.758	.759
.81	.760	.761	.762	.763	.765	.766	.767	.768	.769	.771
.82	.772	.773	.775	.776	.777	.779	.780	.781	.783	.784
.83	.786	.787	.788	.790	.791	.793	.794	.796	.798	.799
.84	.801	.802	.804	.806	.807	.809	.811	.813	.814	.816
.85	.818	.820	.822	.823	.825	.827	.829	.831	.833	.835
.86	.837	.839	.841	.843	.845	.847	.849	.852	.854	.856
.87	.858	.861	.863	.865	.868	.870	.872	.875	.877	.880
.88	.882	.885	.887	.890	.893	.895	.898	.901	.903	.906
.89	.909	.912	.915	.918	.921	.924	.927	.930	.933	.936
.90	.939	.942	.946	.949	.952	.956	.959	.963	.966	.970
.91	.974	.977	.981	.985	.989	.993	.997	1.000	1.004	1.009
.92	1.013	1.017	1.022	1.026	1.031	1.035	1.040	1.044	1.049	1.054
.93	1.059	1.063	1.069	1.074	1.079	1.084	1.090	1.095	1.100	1.107
.94	1.113	1.119	1.125	1.131	1.138	1.144	1.151	1.157	1.164	1.171
.95	1.179	1.186	1.193	1.201	1.209	1.217	1.226	1.235	1.243	1.252
.96	1.261	1.271	1.281	1.291	1.301	1.313	1.324	1.335	1.347	1.360
.97	1.373	1.386	1.400	1.415	1.430	1.446	1.463	1.481	1.499	1.519
.98	1.539	1.561	1.584	1.609	1.636	1.665	1.696	1.730	1.767	1.809
.99	1.855	1.907	1.966	2.036	2.119	2.221	2.352	2.520	2.804	3.339

TABLE X. Lorentz-polarization factor, $(L \cdot p)^2 \left[= \left(\frac{1}{\sin 2\theta} \cdot \frac{1 + \cos^2 2\theta}{2} \right)^2 \right]$, as a function of σ .Valid only for zero level when $\mu = 0$.

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.00	∞	22.42	15.79	12.91	11.21	.30	1.795	1.789	1.783	1.777	1.770
.01	10.01	9.129	8.442	7.908	7.451	.31	1.764	1.758	1.752	1.746	1.740
.02	7.070	6.742	6.453	6.201	5.976	.32	1.734	1.729	1.723	1.717	1.711
.03	5.774	5.588	5.422	5.271	5.128	.33	1.706	1.700	1.695	1.689	1.684
.04	5.000	4.879	4.766	4.660	4.563	.34	1.678	1.673	1.668	1.663	1.657
.05	4.470	4.383	4.301	4.224	4.150	.35	1.652	1.647	1.642	1.637	1.632
.06	4.080	4.014	3.950	3.890	3.831	.36	1.627	1.622	1.617	1.612	1.607
.07	3.776	3.723	3.673	3.624	3.576	.37	1.602	1.598	1.593	1.588	1.584
.08	3.531	3.487	3.445	3.406	3.366	.38	1.579	1.574	1.570	1.566	1.561
.09	3.329	3.291	3.256	3.222	3.189	.39	1.557	1.552	1.548	1.543	1.539
.10	3.157	3.125	3.095	3.065	3.036	.40	1.535	1.531	1.526	1.522	1.518
.11	3.008	2.981	2.955	2.939	2.904	.41	1.514	1.510	1.505	1.501	1.497
.12	2.889	2.855	2.832	2.809	2.787	.42	1.493	1.489	1.485	1.481	1.477
.13	2.765	2.744	2.722	2.702	2.682	.43	1.473	1.470	1.466	1.462	1.458
.14	2.663	2.644	2.623	2.607	2.589	.44	1.454	1.451	1.447	1.443	1.439
.15	2.571	2.554	2.537	2.520	2.504	.45	1.436	1.432	1.429	1.425	1.421
.16	2.488	2.472	2.457	2.442	2.427	.46	1.418	1.414	1.411	1.407	1.404
.17	2.412	2.398	2.384	2.370	2.356	.47	1.400	1.397	1.393	1.390	1.386
.18	2.343	2.330	2.317	2.304	2.291	.48	1.383	1.380	1.377	1.373	1.370
.19	2.279	2.266	2.254	2.243	2.231	.49	1.366	1.363	1.360	1.357	1.354
.20	2.219	2.208	2.197	2.186	2.175	.50	1.350	1.347	1.344	1.341	1.338
.21	2.164	2.154	2.143	2.133	2.123	.51	1.335	1.332	1.329	1.326	1.322
.22	2.113	2.103	2.093	2.084	2.074	.52	1.319	1.316	1.313	1.310	1.307
.23	2.064	2.055	2.046	2.037	2.028	.53	1.304	1.301	1.298	1.295	1.293
.24	2.019	2.011	2.002	1.993	1.985	.54	1.290	1.287	1.284	1.281	1.278
.25	1.977	1.969	1.960	1.952	1.944	.55	1.276	1.273	1.270	1.267	1.264
.26	1.936	1.929	1.921	1.913	1.906	.56	1.261	1.259	1.256	1.253	1.251
.27	1.898	1.891	1.884	1.877	1.869	.57	1.248	1.245	1.243	1.240	1.237
.28	1.862	1.855	1.848	1.842	1.835	.58	1.235	1.232	1.229	1.227	1.224
.29	1.828	1.821	1.815	1.808	1.802	.59	1.221	1.219	1.216	1.214	1.211

TABLE X.—Continued.

	.000	.002	.004	.006	.008		.000	.002	.004	.006	.008
.60	1.209	1.206	1.204	1.201	1.199	1.30	.720	.719	.719	.719	.718
.61	1.196	1.194	1.191	1.189	1.187	1.31	.718	.717	.717	.717	.716
.62	1.184	1.182	1.179	1.177	1.174	1.32	.716	.716	.715	.715	.714
.63	1.172	1.170	1.167	1.165	1.163	1.33	.714	.714	.714	.713	.713
.64	1.160	1.158	1.155	1.153	1.151	1.34	.713	.712	.712	.712	.712
.65	1.148	1.146	1.144	1.142	1.140	1.35	.711	.711	.711	.711	.710
.66	1.138	1.135	1.133	1.130	1.128	1.36	.710	.710	.710	.710	.709
.67	1.126	1.124	1.122	1.120	1.118	1.37	.709	.709	.709	.709	.708
.68	1.115	1.113	1.111	1.109	1.107	1.38	.708	.708	.708	.708	.708
.69	1.105	1.103	1.100	1.098	1.096	1.39	.708	.708	.708	.707	.707
.70	1.094	1.092	1.090	1.088	1.086	1.40	.707	.707	.707	.707	.707
.71	1.084	1.082	1.080	1.078	1.076	1.41	.707	.707	.707	.707	.707
.72	1.074	1.072	1.070	1.068	1.066	1.42	.707	.707	.707	.707	.707
.73	1.064	1.062	1.060	1.058	1.056	1.43	.707	.707	.708	.708	.708
.74	1.054	1.052	1.050	1.048	1.046	1.44	.708	.708	.708	.708	.708
.75	1.045	1.043	1.041	1.039	1.037	1.45	.708	.709	.709	.709	.709
.76	1.035	1.033	1.031	1.029	1.028	1.46	.709	.710	.710	.710	.710
.77	1.026	1.024	1.022	1.020	1.018	1.47	.710	.711	.711	.711	.712
.78	1.016	1.015	1.013	1.011	1.009	1.48	.712	.712	.713	.713	.713
.79	1.007	1.005	1.004	1.002	1.000	1.49	.714	.714	.714	.715	.715
.80	.999	.997	.995	.993	.992	1.50	.715	.716	.716	.717	.717
.81	.990	.988	.986	.985	.983	1.51	.717	.718	.718	.719	.719
.82	.981	.980	.978	.976	.975	1.52	.720	.720	.721	.721	.722
.83	.973	.971	.970	.968	.966	1.53	.723	.723	.724	.724	.725
.84	.965	.963	.961	.960	.958	1.54	.726	.726	.727	.727	.728
.85	.957	.955	.953	.952	.950	1.55	.729	.729	.730	.731	.732
.86	.949	.947	.945	.944	.942	1.56	.732	.733	.734	.735	.735
.87	.941	.939	.937	.936	.934	1.57	.736	.737	.738	.739	.739
.88	.933	.931	.930	.928	.927	1.58	.740	.741	.742	.743	.744
.89	.925	.924	.922	.921	.919	1.59	.745	.746	.747	.747	.748
.90	.918	.916	.915	.913	.912	1.60	.749	.750	.751	.752	.753
.91	.910	.909	.907	.906	.905	1.61	.755	.756	.757	.758	.759
.92	.903	.902	.900	.899	.897	1.62	.760	.761	.762	.763	.765
.93	.896	.895	.893	.892	.890	1.63	.766	.767	.768	.769	.771
.94	.889	.888	.886	.885	.884	1.64	.772	.773	.775	.776	.777
.95	.882	.881	.879	.878	.877	1.65	.779	.780	.781	.783	.784
.96	.875	.874	.873	.871	.870	1.66	.786	.787	.788	.790	.791
.97	.869	.867	.866	.865	.863	1.67	.793	.794	.796	.798	.799
.98	.862	.861	.860	.858	.857	1.68	.801	.802	.804	.806	.807
.99	.856	.855	.853	.852	.851	1.69	.809	.811	.813	.814	.816
1.00	.850	.848	.847	.846	.845	1.70	.818	.820	.822	.823	.825
1.01	.843	.842	.841	.840	.839	1.71	.827	.829	.831	.833	.835
1.02	.837	.836	.835	.834	.833	1.72	.837	.839	.841	.843	.845
1.03	.831	.830	.829	.828	.827	1.73	.847	.849	.852	.854	.856
1.04	.826	.824	.823	.822	.821	1.74	.858	.861	.863	.865	.868
1.05	.820	.819	.818	.817	.815	1.75	.870	.872	.875	.877	.880
1.06	.814	.813	.812	.811	.810	1.76	.882	.885	.887	.890	.893
1.07	.809	.808	.807	.806	.805	1.77	.895	.898	.901	.903	.906
1.08	.803	.802	.801	.800	.799	1.78	.909	.912	.915	.918	.921
1.09	.798	.797	.796	.795	.794	1.79	.924	.927	.930	.933	.936
1.10	.793	.792	.791	.790	.789	1.80	.939	.942	.946	.949	.952
1.11	.788	.787	.786	.785	.784	1.81	.956	.959	.963	.966	.970
1.12	.783	.782	.781	.781	.780	1.82	.974	.977	.981	.985	.989
1.13	.779	.778	.777	.776	.775	1.83	.993	.997	1.000	1.004	1.009
1.14	.774	.773	.772	.771	.771	1.84	1.013	1.017	1.022	1.026	1.031
1.15	.770	.769	.768	.767	.766	1.85	1.035	1.040	1.044	1.049	1.054
1.16	.765	.764	.764	.763	.762	1.86	1.059	1.063	1.069	1.074	1.079
1.17	.761	.760	.760	.759	.758	1.87	1.084	1.090	1.095	1.100	1.107
1.18	.757	.756	.756	.755	.754	1.88	1.113	1.119	1.125	1.131	1.138
1.19	.753	.752	.752	.751	.750	1.89	1.144	1.151	1.157	1.164	1.171
1.20	.749	.749	.748	.747	.747	1.90	1.179	1.186	1.193	1.201	1.209
1.21	.746	.745	.744	.744	.743	1.91	1.217	1.226	1.235	1.243	1.252
1.22	.742	.742	.741	.740	.740	1.92	1.261	1.271	1.281	1.291	1.301
1.23	.739	.738	.738	.737	.736	1.93	1.313	1.324	1.335	1.347	1.360
1.24	.736	.735	.735	.734	.733	1.94	1.373	1.386	1.400	1.415	1.430
1.25	.733	.732	.732	.731	.730	1.95	1.446	1.463	1.481	1.499	1.519
1.26	.730	.729	.729	.728	.728	1.96	1.539	1.561	1.584	1.609	1.636
1.27	.727	.727	.726	.726	.725	1.97	1.665	1.696	1.730	1.767	1.809
1.28	.725	.724	.724	.723	.723	1.98	1.855	1.907	1.966	2.036	2.199
1.29	.722	.722	.721	.721	.720	1.99	2.221	2.352	2.520	2.804	3.339

TABLE XI. ($h^2 + k^2$).

	h																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	0	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000	9.000	10.00	11.00	12.00	13.00	14.00	15.00	0
1	1.000	1.414	2.236	3.162	4.123	5.099	6.083	7.071	8.062	9.055	10.05	11.05	12.04	13.04	14.04	15.03	1
2	2.000	2.236	2.828	3.606	4.472	5.385	6.325	7.280	8.246	9.220	10.20	11.18	12.17	13.15	14.14	15.13	2
3	3.000	3.162	3.606	4.243	5.000	5.831	6.708	7.616	8.544	9.487	10.44	11.40	12.37	13.34	14.32	15.30	3
4	4.000	4.123	4.472	5.000	5.637	6.403	7.211	8.062	8.944	9.849	10.77	11.70	12.65	13.60	14.56	15.52	4
5	5.000	5.099	5.385	5.831	6.403	7.071	7.810	8.602	9.434	10.30	11.18	12.08	13.00	13.93	14.87	15.81	5
6	6.000	6.083	6.325	6.708	7.211	7.810	8.485	9.220	10.00	10.82	11.66	12.53	13.42	14.32	15.23	16.16	6
7	7.000	7.071	7.280	7.616	8.062	8.602	9.220	9.899	10.63	11.40	12.21	13.04	13.89	14.76	15.65	16.55	7
8	8.000	8.062	8.246	8.544	8.944	9.434	10.00	10.63	11.31	12.04	12.81	13.60	14.42	15.26	16.12	17.00	8
9	9.000	9.055	9.220	9.487	9.849	10.30	10.82	11.40	12.04	12.73	13.45	14.21	15.00	15.81	16.64	17.49	9
10	10.00	10.05	10.20	10.44	10.77	11.18	11.66	12.21	12.81	13.45	14.14	14.87	15.62	16.40	17.20	18.03	10
11	11.00	11.05	11.18	11.40	11.70	12.08	12.53	13.04	13.60	14.21	14.87	15.56	16.28	17.03	17.80	18.60	11
12	12.00	12.04	12.17	12.37	12.65	13.00	13.42	13.89	14.42	15.00	15.62	16.28	16.97	17.69	18.44	19.21	12
13	13.00	13.04	13.15	13.34	13.60	13.93	14.32	14.76	15.26	15.81	16.40	17.03	17.69	18.38	19.10	19.85	13
14	14.00	14.04	14.14	14.32	14.56	14.87	15.23	15.65	16.12	16.64	17.20	17.80	18.44	19.10	19.80	20.52	14
15	15.00	15.03	15.13	15.30	15.52	15.81	16.16	16.55	17.00	17.49	18.03	18.60	19.21	19.85	20.52	21.21	15
16	16.00	16.03	16.12	16.28	16.49	16.76	17.09	17.46	17.89	18.36	18.87	19.42	20.00	20.62	21.26	21.93	16
17	17.00	17.03	17.12	17.26	17.46	17.72	18.03	18.38	18.79	19.24	19.72	20.25	20.81	21.40	22.02	22.67	17
18	18.00	18.03	18.11	18.25	18.44	18.68	18.97	19.31	19.70	20.12	20.59	21.10	21.63	22.20	22.80	23.43	18
19	19.00	19.03	19.10	19.24	19.42	19.65	19.92	20.25	20.62	21.02	21.47	21.95	22.47	23.02	23.60	24.21	19
20	20.00	20.02	20.10	20.22	20.40	20.62	20.88	21.19	21.54	21.93	22.36	22.83	23.32	23.85	24.41	25.00	20
21	21.00	21.02	21.10	21.21	21.38	21.59	21.84	22.14	22.47	22.85	23.26	23.71	24.19	24.70	25.24	25.81	21
22	22.00	22.02	22.09	22.20	22.36	22.56	22.80	23.09	23.41	23.77	24.17	24.60	25.06	25.55	26.08	26.63	22
23	23.00	23.02	23.09	23.19	23.35	23.54	23.77	24.04	24.35	24.70	25.08	25.50	25.94	26.42	26.93	27.46	23
24	24.00	24.02	24.08	24.19	24.33	24.52	24.74	25.00	25.30	25.63	26.00	26.40	26.83	27.29	27.78	28.30	24
25	25.00	25.02	25.08	25.18	25.32	25.50	25.71	25.96	26.25	26.57	26.93	27.31	27.73	28.18	28.65	29.15	25
26	26.00	26.02	26.08	26.17	26.31	26.48	26.68	26.93	27.20	27.51	27.86	28.23	28.64	29.07	29.53	30.02	26
27	27.00	27.02	27.07	27.17	27.29	27.46	27.66	27.89	28.16	28.46	28.79	29.15	29.55	29.97	30.41	30.89	27
28	28.00	28.02	28.07	28.16	28.28	28.44	28.64	28.86	29.12	29.41	29.73	30.08	30.46	30.87	31.30	31.76	28
29	29.00	29.02	29.07	29.15	29.27	29.43	29.61	29.83	30.08	30.36	30.68	31.02	31.38	31.78	32.20	32.65	29
30	30.00	30.02	30.07	30.15	30.27	30.41	30.59	30.81	31.05	31.32	31.62	31.95	32.31	32.70	33.11	33.54	30

TABLE XI.—Continued.

		h															
		16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
0	16.00	17.00	18.00	19.00	20.00	21.00	22.00	23.00	24.00	25.00	26.00	27.00	28.00	29.00	30.00	0	
1	16.03	17.03	18.03	19.03	20.02	21.02	22.02	23.02	24.02	25.02	26.02	27.02	28.02	29.02	30.02	1	
2	16.12	17.12	18.11	19.10	20.10	21.10	22.09	23.09	24.08	25.08	26.08	27.07	28.07	29.07	30.07	2	
3	16.28	17.26	18.25	19.24	20.22	21.21	22.20	23.19	24.19	25.18	26.17	27.17	28.16	29.15	30.15	3	
4	16.49	17.46	18.44	19.42	20.40	21.38	22.36	23.35	24.33	25.32	26.31	27.29	28.28	29.27	30.27	4	
5	16.76	17.72	18.68	19.65	20.62	21.59	22.56	23.54	24.52	25.50	26.48	27.46	28.44	29.43	30.41	5	
6	17.09	18.03	18.97	19.92	20.88	21.84	22.80	23.77	24.74	25.71	26.68	27.66	28.64	29.61	30.59	6	
7	17.46	18.38	19.31	20.25	21.19	22.14	23.09	24.04	25.00	25.96	26.93	27.89	28.86	29.83	30.81	7	
8	17.89	18.79	19.70	20.62	21.54	22.47	23.41	24.35	25.30	26.25	27.20	28.16	29.12	30.08	31.05	8	
9	18.36	19.24	20.12	21.02	21.93	22.85	23.77	24.70	25.63	26.57	27.51	28.46	29.41	30.36	31.32	9	
10	18.87	19.72	20.59	21.47	22.36	23.26	24.17	25.08	26.00	26.93	27.86	28.79	29.73	30.68	31.62	10	
11	19.42	20.25	21.10	21.95	22.83	23.71	24.60	25.50	26.40	27.31	28.23	29.15	30.08	31.02	31.95	11	
12	20.00	20.81	21.63	22.47	23.32	24.19	25.06	25.94	26.83	27.73	28.64	29.55	30.46	31.38	32.31	12	
13	20.62	21.40	22.20	23.02	23.85	24.70	25.55	26.42	27.29	28.18	29.07	29.97	30.87	31.78	32.70	13	
14	21.26	22.02	22.80	23.60	24.41	25.24	26.08	26.93	27.78	28.65	29.53	30.41	31.30	32.20	33.11	14	
15	21.93	22.67	23.43	24.21	25.00	25.81	26.63	27.46	28.30	29.15	30.02	30.89	31.76	32.65	33.54	15	
16	22.63	23.35	24.08	24.84	25.61	26.40	27.20	28.02	28.84	29.68	30.53	31.38	32.25	33.12	34.00	16	
17	23.35	24.04	24.76	25.50	26.25	27.02	27.80	28.60	29.41	30.23	31.06	31.91	32.76	33.62	34.48	17	
18	24.08	24.76	25.46	26.17	26.91	27.66	28.43	29.21	30.00	30.81	31.62	32.45	33.29	34.13	34.99	18	
19	24.84	25.50	26.17	26.87	27.59	28.32	29.07	29.83	30.61	31.40	32.20	33.02	33.84	34.67	35.51	19	
20	25.61	26.25	26.91	27.59	28.28	29.00	29.73	30.48	31.24	32.02	32.80	33.60	34.41	35.23	36.06	20	
21	26.40	27.02	27.66	28.32	29.00	29.70	30.41	31.14	31.89	32.65	33.42	34.21	35.00	35.81	36.62	21	
22	27.20	27.80	28.43	29.07	29.73	30.41	31.11	31.83	32.56	33.30	34.06	34.83	35.61	36.40	37.20	22	
23	28.02	28.60	29.21	29.83	30.48	31.14	31.83	32.53	33.24	33.97	34.71	35.47	36.24	37.01	37.80	23	
24	28.84	29.41	30.00	30.61	31.24	31.89	32.56	33.24	33.94	34.66	35.38	36.12	36.88	37.64	38.42	24	
25	29.68	30.23	30.81	31.40	32.02	32.65	33.30	33.97	34.66	35.36	36.07	36.80	37.54	38.29	39.05	25	
26	30.53	31.06	31.62	32.20	32.80	33.42	34.06	34.71	35.38	36.07	36.77	37.48	38.21	38.95	39.70	26	
27	31.38	31.91	32.45	33.02	33.60	34.21	34.83	35.47	36.12	36.80	37.48	38.18	38.90	39.62	40.36	27	
28	32.25	32.76	33.29	33.84	34.41	35.00	35.61	36.24	36.88	37.54	38.21	38.90	39.60	40.31	41.04	28	
29	33.12	33.62	34.13	34.67	35.23	35.81	36.40	37.01	37.64	38.29	38.95	39.62	40.31	41.01	41.73	29	
30	34.00	34.48	34.99	35.51	36.06	36.62	37.20	37.80	38.42	39.05	39.70	40.36	41.04	41.73	42.43	30	

TABLE XII. $(h^2 + k^2 + hk)^{1/2}$.

	h																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	0	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000	9.000	10.00	11.00	12.00	13.00	14.00	15.00	0
1	1.000	1.732	2.646	3.606	4.583	5.568	6.557	7.550	8.544	9.539	10.54	11.53	12.53	13.53	14.53	15.52	1
2	2.000	2.646	3.606	4.359	5.292	6.245	7.211	8.185	9.165	10.15	11.14	12.12	13.11	14.11	15.10	16.09	2
3	3.000	3.606	4.359	5.196	6.083	7.000	7.937	8.888	9.849	10.82	11.79	12.77	13.75	14.73	15.72	16.70	3
4	4.000	4.583	5.292	6.083	6.928	7.810	8.718	9.644	10.58	11.53	12.49	13.45	14.42	15.39	16.37	17.35	4
5	5.000	5.568	6.245	7.000	7.810	8.660	9.539	10.44	11.36	12.29	13.23	14.18	15.13	16.09	17.06	18.03	5
6	6.000	6.557	7.211	7.937	8.718	9.539	10.39	11.27	12.17	13.08	14.00	14.93	15.87	16.82	17.78	18.73	6
7	7.000	7.550	8.185	8.888	9.644	10.44	11.27	12.12	13.00	13.89	14.80	15.72	16.64	17.58	18.52	19.47	7
8	8.000	8.544	9.165	9.849	10.58	11.36	12.17	13.00	13.86	14.73	15.62	16.52	17.44	18.36	19.29	20.22	8
9	9.000	9.539	10.15	10.82	11.53	12.29	13.08	13.89	14.73	15.59	16.46	17.35	18.25	19.16	20.07	21.00	9
10	10.00	10.54	11.14	11.79	12.49	13.23	14.00	14.80	15.62	16.46	17.32	18.19	19.08	19.97	20.88	21.79	10
11	11.00	11.53	12.12	12.77	13.45	14.18	14.93	15.72	16.52	17.35	18.19	19.05	19.92	20.81	21.70	22.61	11
12	12.00	12.53	13.11	13.75	14.42	15.13	15.87	16.64	17.44	18.25	19.08	19.92	20.78	21.66	22.54	23.43	12
13	13.00	13.53	14.11	14.73	15.39	16.09	16.82	17.58	18.36	19.16	19.97	20.81	21.66	22.52	23.39	24.27	13
14	14.00	14.53	15.10	15.72	16.37	17.06	17.78	18.52	19.29	20.07	20.88	21.70	22.54	23.39	24.25	25.12	14
15	15.00	15.52	16.09	16.70	17.35	18.03	18.73	19.47	20.22	21.00	21.79	22.61	23.43	24.27	25.12	25.98	15
16	16.00	16.52	17.09	17.69	18.33	19.00	19.70	20.42	21.17	21.93	22.72	23.52	24.33	25.16	26.00	26.85	16
17	17.00	17.52	18.08	18.68	19.31	19.97	20.66	21.38	22.11	22.87	23.64	24.43	25.24	26.06	26.89	27.73	17
18	18.00	18.52	19.08	19.67	20.30	20.95	21.63	22.34	23.07	23.81	24.58	25.36	26.15	26.96	27.78	28.62	18
19	19.00	19.52	20.07	20.66	21.28	21.93	22.61	23.30	24.02	24.76	25.51	26.29	27.07	27.87	28.69	29.51	19
20	20.00	20.52	21.07	21.66	22.27	22.91	23.58	24.27	24.98	25.71	26.46	27.22	28.00	28.79	29.60	30.41	20
21	21.00	21.52	22.07	22.65	23.26	23.90	24.56	25.24	25.94	26.66	27.40	28.16	28.93	29.72	30.51	31.32	21
22	22.00	22.52	23.07	23.64	24.25	24.88	25.53	26.21	26.91	27.62	28.35	29.10	29.87	30.64	31.43	32.23	22
23	23.00	23.52	24.06	24.64	25.24	25.87	26.51	27.18	27.87	28.58	29.31	30.05	30.81	31.58	32.36	33.15	23
24	24.00	24.52	25.06	25.63	26.23	26.85	27.50	28.16	28.84	29.55	30.27	31.00	31.75	32.51	33.29	34.07	24
25	25.00	25.51	26.06	26.63	27.22	27.84	28.48	29.14	29.82	30.51	31.22	31.95	32.70	33.45	34.22	35.00	25
26	26.00	26.51	27.06	27.62	28.21	28.83	29.46	30.12	30.79	31.48	32.19	32.91	33.65	34.39	35.16	35.93	26
27	27.00	27.51	28.05	28.62	29.21	29.82	30.45	31.10	31.76	32.45	33.15	33.87	34.60	35.34	36.10	36.86	27
28	28.00	28.51	29.05	29.61	30.20	30.81	31.43	32.08	32.74	33.42	34.12	34.83	35.55	36.29	37.04	37.80	28
29	29.00	29.51	30.05	30.61	31.19	31.80	32.42	33.06	33.72	34.39	35.09	35.79	36.51	37.24	37.99	38.74	29
30	30.00	30.51	31.05	31.61	32.19	32.79	33.41	34.04	34.70	35.37	36.06	36.76	37.47	38.20	38.94	39.69	30

TABLE XII.—Continued.

	h															
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
0	16.00	17.00	18.00	19.00	20.00	21.00	22.00	23.00	24.00	25.00	26.00	27.00	28.00	29.00	30.00	0
1	16.52	17.52	18.52	19.52	20.52	21.52	22.52	23.52	24.52	25.51	26.51	27.51	28.51	29.51	30.51	1
2	17.09	18.08	19.08	20.07	21.07	22.07	23.07	24.06	25.06	26.06	27.06	28.05	29.05	30.05	31.05	2
3	17.69	18.68	19.67	20.66	21.66	22.65	23.64	24.64	25.63	26.63	27.62	28.62	29.61	30.61	31.61	3
4	18.33	19.31	20.30	21.28	22.27	23.26	24.25	25.24	26.23	27.22	28.21	29.21	30.20	31.19	32.19	4
5	19.00	19.97	20.95	21.93	22.91	23.90	24.88	25.87	26.85	27.84	28.83	29.82	30.81	31.80	32.79	5
6	19.70	20.66	21.63	22.61	23.58	24.56	25.53	26.51	27.50	28.48	29.46	30.45	31.43	32.42	33.41	6
7	20.42	21.38	22.34	23.30	24.27	25.24	26.21	27.18	28.16	29.14	30.12	31.10	32.08	33.06	34.04	7
8	21.17	22.11	23.07	24.02	24.98	25.94	26.91	27.87	28.84	29.82	30.79	31.76	32.74	33.72	34.70	8
9	21.93	22.87	23.81	24.76	25.71	26.66	27.62	28.58	29.55	30.51	31.48	32.45	33.42	34.39	35.37	9
10	22.72	23.64	24.58	25.51	26.46	27.40	28.35	29.31	30.27	31.22	32.19	33.15	34.12	35.09	36.06	10
11	23.52	24.43	25.36	26.29	27.22	28.16	29.10	30.05	31.00	31.95	32.91	33.87	34.83	35.79	36.76	11
12	24.33	25.24	26.15	27.07	28.00	28.93	29.87	30.81	31.75	32.70	33.65	34.60	35.55	36.51	37.47	12
13	25.16	26.06	26.96	27.87	28.79	29.72	30.64	31.58	32.51	33.45	34.39	35.34	36.29	37.24	38.20	13
14	26.00	26.89	27.78	28.69	29.60	30.51	31.43	32.36	33.29	34.22	35.16	36.10	37.04	37.99	38.94	14
15	26.85	27.73	28.62	29.51	30.41	31.32	32.23	33.15	34.07	35.00	35.93	36.86	37.80	38.74	39.69	15
16	27.71	28.58	29.46	30.35	31.24	32.14	33.05	33.96	34.87	35.79	36.72	37.64	38.57	39.51	40.45	16
17	28.58	29.44	30.32	31.19	32.08	32.97	33.87	34.77	35.68	36.59	37.51	38.43	39.36	40.29	41.22	17
18	29.46	30.32	31.18	32.05	32.92	33.81	34.70	35.59	36.50	37.40	38.31	39.23	40.15	41.07	42.00	18
19	30.35	31.19	32.05	32.91	33.78	34.66	35.54	36.43	37.32	38.22	39.13	40.04	40.95	41.87	42.79	19
20	31.24	32.08	32.92	33.78	34.64	35.51	36.39	37.27	38.16	39.05	39.95	40.85	41.76	42.67	43.59	20
21	32.14	32.97	33.81	34.66	35.51	36.37	37.24	38.12	39.00	39.89	40.78	41.68	42.58	43.49	44.40	21
22	33.05	33.87	34.70	35.54	36.39	37.24	38.11	38.97	39.85	40.73	41.62	42.51	43.41	44.31	45.21	22
23	33.96	34.77	35.59	36.43	37.27	38.12	38.97	39.84	40.71	41.58	42.46	43.35	44.24	45.13	46.03	23
24	34.87	35.68	36.50	37.32	38.16	39.00	39.85	40.71	41.57	42.44	43.31	44.19	45.08	45.97	46.86	24
25	35.79	36.59	37.40	38.22	39.05	39.89	40.73	41.58	42.44	43.30	44.17	45.04	45.92	46.81	47.70	25
26	36.72	37.51	38.31	39.13	39.95	40.78	41.62	42.46	43.31	44.17	45.03	45.90	46.78	47.66	48.54	26
27	37.64	38.43	39.23	40.04	40.85	41.68	42.51	43.35	44.19	45.04	45.90	46.77	47.63	48.51	49.39	27
28	38.57	39.36	40.15	40.95	41.76	42.58	43.41	44.24	45.08	45.92	46.78	47.63	48.50	49.37	50.24	28
29	39.51	40.29	41.07	41.87	42.67	43.49	44.31	45.13	45.97	46.81	47.66	48.51	49.37	50.23	51.10	29
30	40.45	41.22	42.00	42.79	43.59	44.40	45.21	46.03	46.86	47.70	48.54	49.39	50.24	51.10	51.96	30

Further Improvement in the Resolving Power of the Electron Microscope

JAMES HILLIER

RCA Laboratories, Princeton, New Jersey

(Received February 2, 1946)

THERE are a number of independent possible methods of determining experimentally the electron optical properties of the magnetic electron microscope objectives in use at the present time.¹ From these determinations it has been possible to obtain corresponding independent calculated values for the maximum resolving power which should be attainable with the magnetic electron microscope. All such values have been found to lie in the range 5A to 10A. On the other hand, direct experimental determinations of the resolving powers of existing instruments have given no values below 20A. Furthermore, publications reporting systematic investigations of the reasons for this discrepancy have been singularly lacking.

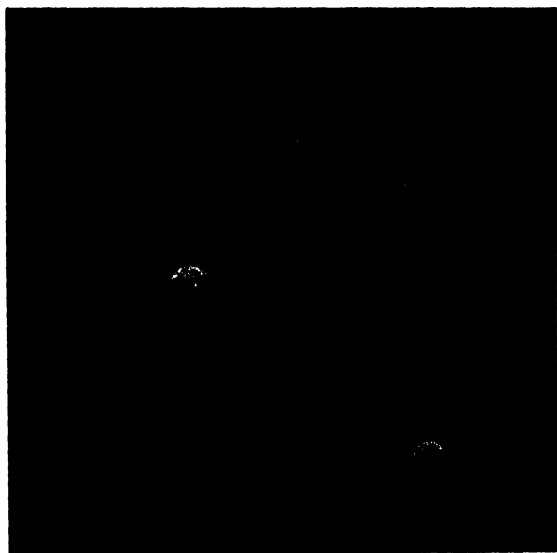


FIG. 1a. An electron micrograph of polystyrene latex on a collodion plus silica substrate and shadowed with approximately 10A of chromium. Additional information: Electronic magnification, 14,600 $\times$. Focal length of the objective 2.5 mm. No limiting aperture. Angular aperture of illumination 10^{-4} rad., approximately. Accelerating potential 50 kv. Recorded on medium lantern slide; 5-second exposure. Developed 3 minutes in D-72. Micrograph obtained by R. F. Baker. Specimen prepared by N. R. Davidson and R. F. Baker.

¹ V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, New York, 1945).

Research workers interested in the problem of improving the ultimate performance of the electron microscope have been well aware of the absence of serious work on this important problem. They have also realized, however, the futility of attempting to carry out a systematic study when significant results are obtainable in only an extremely small percentage of the experiments. For instance, in the author's own laboratory, only 10–15 exposures definitely showing a resolving power of 20A had been obtained from over 25,000 exposures made during the five-year period 1940–1945.

During this time, the efforts of electron microscope designers and engineers had been concentrated on the elimination of all those defects which provided accidental disturbances of the images. Recently, this work was greatly accelerated in this laboratory by the development of a new gun providing a twenty-fold increase in the intensity of the image and considerable

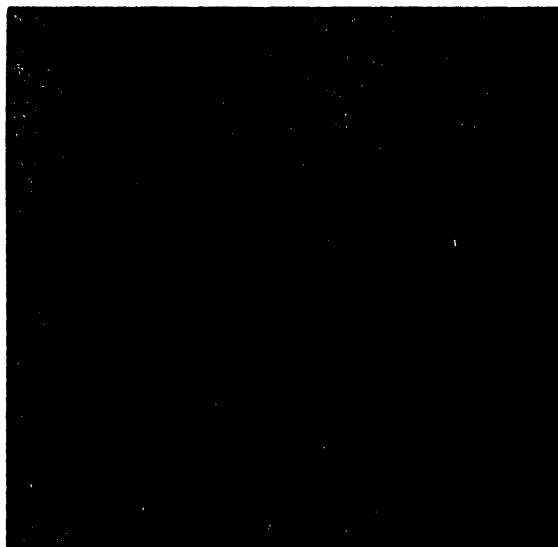


FIG. 1b. A 16 $\times$ enlargement of a portion of Fig. 1a. An attempt has been made to eliminate some of the effects of photographic grain by slightly defocusing the enlarger. Chromium particles can be distinguished from photographic grain by the fact that they are elongated as a result of a slight astigmatism of the objective.



FIG. 2. A photometric trace of the sharp edge of one of the particles of polystyrene latex shown in Fig. 1a.

improvement in the electron optical properties of the specimen illumination.^{2,3} This improvement has made practical the use of a telescopic viewing device (10 $\times$) similar to that described by von Ardenne,⁴ which enables the operator to examine the final image visually at magnifications as high as 300,000 $\times$ at satisfactory intensities and under the best conditions for making the final photographic exposures.

The accumulation of electron microscope engineering knowledge and the above mentioned developments have resulted in an instrument in this laboratory (a modified RCA Model EMU) which when operated by the author or his associates seldom fails to provide images showing resolutions of the order of 20A (measured center to center on definitely resolved particles).

With this degree of performance available it was soon very obvious that it has been a *lack of axial symmetry* in the magnetic fields of the lenses that has limited their resolving power. The author has since undertaken to produce a lens possessing a greater degree of symmetry and at the time of writing has obtained at least one lens which shows considerable improve-

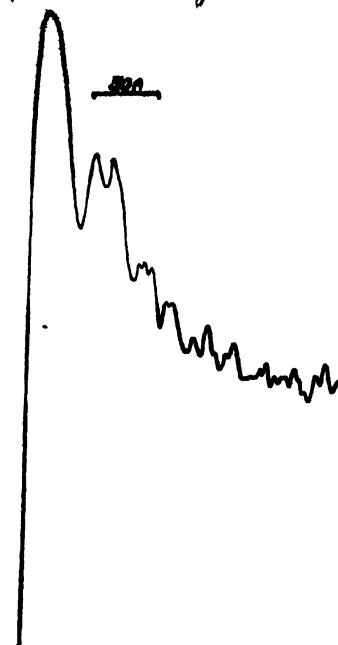


FIG. 3. A photometric trace of the Fresnel fringes obtained at an edge in a very slightly out of focus image of MgO.

ment. Since the techniques used are exploratory in nature and the accuracy achieved is not yet susceptible to measurement, neither will be discussed here. The purpose of this note is rather to report that a few images have been obtained with this lens which show a resolving power better than 13A and possibly as low as 8.5A corresponding to a useful magnification of over 200,000!

In order to obtain a representative value for the limiting resolving power, three different measurements were used in its evaluation: (1) The distance between centers of just resolved particles—13A (see Fig. 1). (2) Half the width of the contour in the image of a sharp edge, taken from external maximum to the internal minimum—8.5A (see Fig. 2). (3) The estimated minimum possible separation, consistent with resolution, between Fresnel fringes in the image of a sharp edge—10A (see Fig. 3). These measurements give a mean value of 10.5A for the resolving power of the present lens.

This work indicates that the calculated limiting resolving power of the present magnetic lenses is obtainable in practice. On the other hand, it has revealed an imposing list of technical problems which must be solved before such

² J. Hillier and R. F. Baker, "A discussion of the illuminating system of the electron microscope," *J. App. Phys.* **16**, 483 (1945).

³ J. Hillier, and R. F. Baker, "On the improvement of resolution in the electron diffraction cameras," *J. App. Phys.* **17**, 12 (1946).

⁴ M. von Ardenne, *Elektronen Übermikroskopie* (Verlagsbuchhandlung Julius Springer, Berlin, 1940).

high resolving power will be consistently obtainable or before it will be available for general scientific research.

Through suggestions and discussions a number of workers have contributed to this development.

It is with pleasure that the author acknowledges the part played in the work by his co-workers in the RCA Laboratories in Princeton and by Mr. Perry C. Smith and his associates of the RCA Victor Division in Camden.

Preparation of Electron Microscope Specimens for Determination of Particle Size Distribution in Aqueous Suspensions

A. M. CRAVATH, A. E. SMITH, J. R. VINOGRAD, AND J. N. WILSON
Shell Development Company, Emeryville, California

(Received December 17, 1945)

Small droplets of the suspension to be examined are deposited from a mist upon a specimen film whose surface has been treated to promote wetting by the droplets. This procedure makes the entire residue left by a whole droplet available for examination and promotes uniform dispersion of the deposit.

IN order to overcome difficulties experienced in measuring the size distribution of particles in polystyrene latices the following procedures for mounting specimens for electron microscope examination were developed. It is believed they may be of general utility.

When diluted latex (*ca.* 0.01 percent polystyrene) was applied with a medicine dropper to untreated cellulose ester specimen films and allowed to dry, there appeared on the films some areas in which the particles were well dispersed, but these areas were adjacent to dense streaks and ridges of particles which could not be measured. It was not certain that the two types of deposit had identical compositions. When the diluted latex was applied as a fine mist, the whole deposit of some droplets was well dispersed but other droplets gave deposits having a dense ring of unmeasurable particles. With certain samples good deposits were rare. It seemed probable that the rings were caused by poor wetting, the liquid surface meeting the film at a large contact angle and pushing the particles ahead as it receded until such a high concentration was reached that a ridge was deposited. This supposition was supported by the observation that droplets of visible size maintained large contact angles when they were made to shrink by blotting with the edge of a filter paper. A few experiments with Aerosol OT indicated that an effective concentration of wetting agent would contaminate the samples excessively.

The above difficulties were overcome by (a) hydrolyzing the surface of the cellulose ester film enough to promote good wetting and uniform deposition of droplets and (b) applying the latex in the form of extremely fine droplets.

A convenient procedure is to place a drop of ten percent sodium hydroxide solution on a cellulose



FIG. 1. Polystyrene latex. Good dispersion from mist droplet on untreated film. 21,000X.

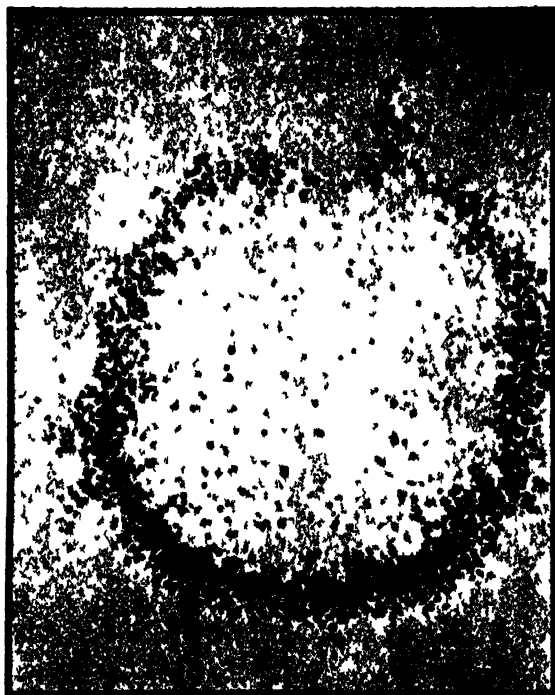


FIG. 2. Polystyrene latex Poor dispersion from mist droplet on untreated film 11,000 $\times$



FIG. 3. Polystyrene latex Good dispersion from mist droplet on treated film 13,000 $\times$.

acetate-butyrate film after the film is attached to the specimen screen. The drop should not reach quite to the edge of the film. The screen is held up to the light and watched. After a few seconds a hole sometimes appears in the film, usually near the edge, and gradually grows. If the hole reaches the center before the treatment has proceeded for the time that experience shows necessary to make the film wettable (about 20 seconds), the specimen is discarded. Satisfactory films are rinsed with distilled water from a wash bottle and dried. The film is placed on a microscope slide in a horizontal test tube (ca. 3×10 cm), and the aqueous sample is sprayed from an atomizer which gives a fine mist (e.g., DeVilbiss No. 14), the spray being directed against the end of the tube so that the larger droplets deposit there and only the finer droplets travel back toward the mouth and settle on the film. (The diameter of the droplet which left the deposit shown in Fig. 1 has been calculated from the solid content, the number of particles and the size distribution to have been 7.4 microns. Since the diameter of the deposit is about 3 microns, considerable evapora-

tion and shrinkage must have occurred before the particles were deposited.)

The surface hydrolysis procedure was found to reduce the contact angle between the film and a visible water droplet to approximately zero.

An alternative procedure for surface hydrolysis has been tried in which the cellulose ester films are cast on sodium hydroxide solution instead of on water. Solutions of cellulose acetate-butyrate or nitrate in amyl acetate were found not to spread on the surface of ten percent sodium hydroxide, but spreading did occur on a three percent hydroxide solution. Nitrocellulose films were attacked so rapidly that they could not be transferred to the specimen screens. Cellulose acetate-butyrate films rapidly grew brittle and weak and soon took on a dull appearance, although they were still strong enough to be transferred to the specimen screen several minutes after forming. In the electron microscope, specimens prepared with these films were indistinguishable from those prepared by the more convenient method described first.

Figures 1-3 show examples of deposits on treated and untreated film.

Here and There

New Appointments

Dr. John D. Strong joined the Johns Hopkins University staff in February as professor of physics.

Dr. William V. Houston, formerly professor of physics at California Institute of Technology, began his new work as president of Rice Institute on March 1.

Dr. Frank G. Dunnington, who has been at the Radiation Laboratory, Massachusetts Institute of Technology, during the war, has returned to active duty at Rutgers University and has been promoted to associate professor. He will become chairman of the Physics Department on the retirement this year of Dr. George Winchester, present department head.

Dr. Raymond G. Spencer, formerly chairman of the metals and minerals research division of the Armour Research Foundation, Chicago, is now director of the Washington University Foundation, St. Louis.

The following persons were recently added to the staff of Philips Laboratories, Inc., Irvington, New York: Dr. James G. Black, Carol M. Veronda, Frank Grace, and George A. Espersen. Dr. Victor Wouk has joined the Engineering Laboratories staff of North American Philips Company, Inc., at Dobbs Ferry, New York.

Dr. George E. Ziegler, Executive Scientist of the Midwest Research Institute, Kansas City, Missouri, has been elected chairman and editor of the recently established Technical Societies Council of the Kansas City Area. He was formerly chairman of physics research at Armour Research Foundation in Chicago.

Westinghouse Science Writing Award Fund

The American Association for the Advancement of Science has announced the establishment of the George Westinghouse Science Writing Award Fund to give national recognition to newspaper writers and newspapers contributing most to better popular understanding of the achievements of science and technology. The fund has been provided by the Westinghouse Educational Foundation in commemoration of the 100th anniversary of the birth of George Westinghouse. It provides an annual cash award of \$1000 to a newspaper writer for outstanding science reporting of the year, and a citation to the newspaper whose science news coverage in the preceding year is adjudged most complete and authoritative and most interestingly presented. The first annual awards, covering the year 1946, will be made at the winter meeting of the A.A.A.S. in December 1946.

American Chemical Society Fellowships

The American Chemical Society will award fellowships aggregating \$210,000 in 1946 and 1947 to aid in the training of chemists and chemical engineers. Predoctoral students whose work was interrupted by the war will receive \$100,000. Grants of \$110,000 will be made to holders of the doctor's degree who desire to devote one or two years to research and teaching. The predoctoral fellowships will carry an annual stipend of \$1200 for

single candidates and \$1800 for married candidates, plus a maximum of \$500 a year for tuition and laboratory charges. In the selection of recipients, great weight will be given to each applicant's sacrifices and contributions to the war effort, with special consideration for those who had begun their graduate work before they entered war service.

"The granting of these predoctoral fellowships is an immediate post-war measure," the Society announced. "While successful candidates may reasonably expect continuation of support until their postgraduate work is complete (for a maximum period of three years), continued aid will be contingent upon satisfactory progress by the fellow."

Postdoctoral fellowships will carry a stipend of \$2500 a year, with the understanding that the institutions at which the recipients study will provide an additional sum for teaching. Awards will be for one year, and will be renewable for a second year upon proof of satisfactory performance. Chemists and chemical engineers who received the doctor's degree since January 1, 1940, and have been engaged in war work, or persons who will have obtained the doctorate during the twelve-month period prior to November 1 of the year in which the award is made, are eligible for postdoctoral fellowships. Applications should be sent to Dr. W. Albert Noyes, Jr., head of the Chemistry Department, University of Rochester, and must be submitted prior to February 1, 1947, by those who intend to begin work by November 1, 1947. Application blanks may be obtained from Dr. Noyes or from the Society's headquarters, 1155—16th Street, N.W., Washington 6, D. C.

Officers for Division of High Polymer Physics

New officers for the Division of High Polymer Physics of the American Physical Society for 1946 are as follows:

Chairman: Howard A. Robinson, Armstrong Cork Company

Vice Chairman: S. L. Gerhard, U. S. Rubber Company

Secretary: W. James Lyons, Firestone Tire and Rubber Company

Treasurer: Lawrence A. Wood, National Bureau of Standards

Adolph Lomb Medal Notice

The Adolph Lomb Medal Committee of the Optical Society of America is engaged in helping to name a recipient for its award which should be made "to a person under thirty (30) years of age who shall have made a noteworthy contribution to optics" (J. Opt. Soc. Am., June, 1940). The choice this year is made more difficult by the wartime restrictions and lack of publications on optical developments of the last several years. To help the committee avoid making a glaring omission, if you know of a possible candidate for this award, would you please communicate your suggestion together with a brief presentation of his or her achievements to some member of the Committee, preferably the chairman, *BEFORE JUNE 1, 1946*? Members of the Committee are:

Dr. Selig Hecht, Columbia University, New York, New York

Dr. Elmer Hutchingson, Case School of Applied Science, Cleveland, Ohio

Dr. W. F. Meggers, National Bureau of Standards, Washington, D. C.

Dr. George S. Monk, University of Chicago, Chicago, Illinois

Dr. Brian O'Brien, University of Rochester, Rochester, New York

Chairman: Dr. Mary E. Wurga, University of Pittsburgh, Pittsburgh 13, Pennsylvania

New Books

Sequential Analysis of Statistical Data: Applications

Prepared by the STATISTICAL RESEARCH GROUP OF COLUMBIA UNIVERSITY for The Applied Mathematics Panel, National Defense Research Committee, Office of Scientific Research and Development. Columbia University Press, New York, 1945. Price \$6.25

One of the developments fostered by wartime shortages of time and man-power is the sequential method of testing a statistical hypothesis. It differs from standard sampling plans in that the sample size is a random variable, testing being stopped whenever a decision is indicated at a pre-set level of assurance. As each unit is inspected, the data accumulated are analyzed and one of three alternatives is indicated: accept hypothesis, reject hypothesis, or continue testing. This plan results in termination of testing at the earliest possible time, reductions in the amount of data required for the same risks of wrong decisions frequently running as high as 50 percent compared with constant-sample-size plans.

Against this advantage must be weighed the disadvantages in industrial inspection of increased responsibility on the inspector and difficulty of guarding against inspector bias as a decision is approached. Nevertheless, the method has been used successfully in many plants for acceptance of war material, and the disadvantages mentioned are usually not serious in applications to experimentation.

The book reviewed here is a manual in which the design of a sequential plan is reduced to the solution of simple algebraic equations followed by a graphical or tabular scoring of results. It is a revised edition of the restricted report originally prepared by H. A. Freeman from the theory developed by A. Wald in 1943, and includes more recent theoretical developments. The form of the original report is retained—six self-contained sections and two appendices, all in pamphlet form and gathered in a loose-leaf ring binder. Each of the sections (except the first, which is a general introduction) takes up a particular type of problem, presents the equations necessary for setting up a test plan, and works a typical example. Thus sequential analysis is applied to data from inspection by attributes (binomial distribution); to inspection by variables (normal distribution), where either the mean or the variability is questioned; and to comparison of data with a standard value, or with data from an alternative method or process. The appendices discuss the nature and theory of sequential analysis, although, for greater mathematical rigor, the reader is referred to the works of A. Wald.

In order to capitalize on the ease with which it may be removed from the binder, each section is made independent of the others by repetition of an adequate explanation of terminology and ideas. The book should prove of considerable value to those concerned with analysis of statistical data in experimentation or acceptance sampling.

D. K. WRIGHT, JR.
Case School of Applied Science

Basic Mathematics for War and Industry

BY PAUL H. DAUS, JOHN M. GLEASON, AND WILLIAM H. WHYBURN. Pp. 11+277, The Macmillan Company, New York, 1944. Price \$2.00.

This book consists of five chapters treating, in that order, arithmetic, algebra, geometry, plane trigonometry, and solid geometry and spherical trigonometry. Three tables are appended—of powers and roots, of four place common logarithms, and of four place trigonometric functions (both natural and logarithmic values). Answers are provided for every problem. A protractor, graduated both in degrees and mils, is inserted in the back cover.

The chapter on arithmetic treats the elementary operations, on integers and fractions, common and decimal, approximation and significant figures, percentage and extraction of square roots. Under algebra, the book deals with the fundamental operations on polynomials and fractions (for some reason, division of polynomials is omitted), factoring, ratio and proportion, radicals (square root only), linear equations and systems of linear equations, quadratic equations (one misses systems of quadratic equations), and graphical solutions of equations. The geometry of the triangle, polygon, and circle is taken up in the third chapter. Here the authors deftly dispose of the difficulties that greet the student in the conventional text by frankly presenting as axioms such theorems as those on the congruence of triangles under various conditions or the proportionality of sides for triangles with equal angles. The chapter is enriched by sections treating the triangle of velocity in air navigation, the wind star, radius of action, interception, angular and rim speeds. In the work on trigonometry, the authors make efficient use of the line values of the trigonometric functions, manage to depart from the conventional texts by presenting what they call the cotangent law and the altitude law for the triangle, treat thoroughly the solution of the triangles, including the aid of logarithms, and enliven it all by presenting, both in the text and in the exercises, genuinely practical and stimulating problems. The last chapter presents the rudiments of solid geometry in fifteen brief, but well done, pages (the various theorems dealing with mensuration, in solid as well as plane geometry, are disposed of as formulas, stated in the chapter on algebra, and given without proof), the rest of the chapter being given over to a good treatment of spherical trigonometry, including the celestial sphere, and once again, such a beguiling topic as "mounting of parts on a tilting table for machining."

Throughout, the exposition is fresh and in many places original, the applications to shop and technical problems numerous, natural, and attractive, and the figures are well drawn. In all, the authors have succeeded in packing a surprisingly large amount of material, and organizing that material well, in the brief space of this book. The authors' purpose, stated in the preface, to provide a useful compendium of elementary mathematics for technical and shop workers, has been achieved here with substantial adequacy.

M. MORGAN
Case School of Applied Science

Journal of Applied Physics

Volume 17, Number 5

May, 1946

The Control of Atomic Energy

DURING the past few months much progress has been made in the formulation of a type of control of atomic energy which, if adopted, would not only make it difficult for atomic bombs to be used as weapons of war, but would also encourage and stimulate the beneficial application of atomic energy in science, medicine, and industry. The progress that has been made is remarkable not only in its extent but also in the large number of individuals who have taken part in it. It is reported that over 70,000 letters have been written to the special Senate Committee on Atomic Energy headed by Senator Brien McMahon. Great credit for this activity should go to the Atomic Scientists of Chicago and their bi-monthly Bulletin,¹ to the striking report to the public on the meaning of the atomic bomb entitled *One World or None*,² and the very extensive public appearances of the leading scientists in this field.

One concrete result of this aroused public interest is the unanimous approval by the McMahon committee of a bill which sets up for the domestic control of atomic energy a full-time Commission consisting of five civilians, with three supplementary committees which may act as checks. These are (1) a military liaison committee; (2) a general advisory committee consisting of nine civilians; and (3) a joint Congressional committee consisting of nine members. The Commission will assume title to all sources of fissionable material, all such materials, and all facilities for developing them, but will allow the use of fissionable material by industry, individuals, and medical and educational institutions. The effect of this bill on the dissemination of scientific information is not absolutely clear, and close scrutiny of the bill

from this point of view will be necessary when it is reported out of the Committee.

The most important of the recent publications on the international control of atomic energy is the report of the Secretary of State's Committee on Atomic Energy.³ In an extremely fascinating and lucid style this report describes how a group of distinguished consultants studying the control problem for the Committee changed their outlook from one of deep pessimism to optimism as their study progressed. There gradually evolved what seems like a workable scheme which now carries the full endorsement of the Committee. An essential feature of the scheme is that an international Atomic Development Authority would be established which would take the lead in developing peacetime applications of atomic energy as well as performing an inspection service to reduce the possibility of surreptitious activities which may lead to atomic warfare. The plan divides activities in the field of atomic energy into "safe" and "dangerous." The "dangerous" consist chiefly in the provision of raw materials, quantity production of fissionable materials, and use of these materials in atomic weapons. The "safe" consist of almost all other applications in science, medicine, and technology in which U235 and plutonium can be used in a denatured state without loss of effectiveness. The proposal is that the dangerous activities be carried on only by the international Authority, and the safe by interested individuals or groups in any nation. Thus the inspection service would not be prohibitive, and peaceful applications would be stimulated rather than retarded. The reading of this important report by all physicists is highly recommended.—E. H.

¹ *Bulletin of the Atomic Scientists*, Office 1126 East 59 Street, Chicago, Illinois. Subscription \$2 a year.

² *One World or None* (Whittlesey House, McGraw-Hill Book Company, Inc., New York, 1946). \$1.

³ *A Report on the International Control of Atomic Energy*, Department of State Publication 2498. For sale by Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. Price 20 cents.

Special Issue on High Polymers

All of the papers in this issue were presented at the second regular meeting of the Division of High Polymer Physics of the American Physical Society at Columbia University, January 24-26, 1946.

Elastic Properties of Cork

I. Stress Relaxation of Compressed Cork

S. L. DART AND EUGENE GUTH

Polymer Physics Laboratory, University of Notre Dame, Notre Dame, Indiana

The stress relaxation of corky materials is of theoretical significance for the understanding of such two-phase (corky material+air) systems, and is of great practical interest, since these materials are commonly used as gaskets, seals, etc. Stress relaxation was studied at various degrees of compression and at temperatures ranging from 30°C to 200°C. The experimental data were obtained from an automatic stress-relaxation machine which employed the principle of a chainomatic balance. It was found that the stress-time curves for various compressions could be obtained from each other by multiplication. This experimental fact is generalized by the statement that the stress

is a product of two functions. The first factor is essentially identical with the S-shaped load-compression curve of cork and is independent of time. The second factor is shown to be a linear function of $\log t$ and is independent of compression. The linear stress-log time curve is found to persist over a wide range of times and to continue until the stress has decayed to zero at the higher temperatures. The decay time as obtained either by actual experiment or by extrapolation is shown as a function of temperature. This gives a fairly complete picture of the stress-time-temperature behavior of cork under compression.

INTRODUCTION

IT is surprising how little the average person knows about cork even though he uses it every day in his car as gaskets, vibration insulation, and in many other ways, such as closures, stoppers, heat insulation, etc.

Since the literature on cork, especially the scientific literature, is quite meager, it will not be out of place to give a brief outline of what cork really is before presenting any detailed work.*

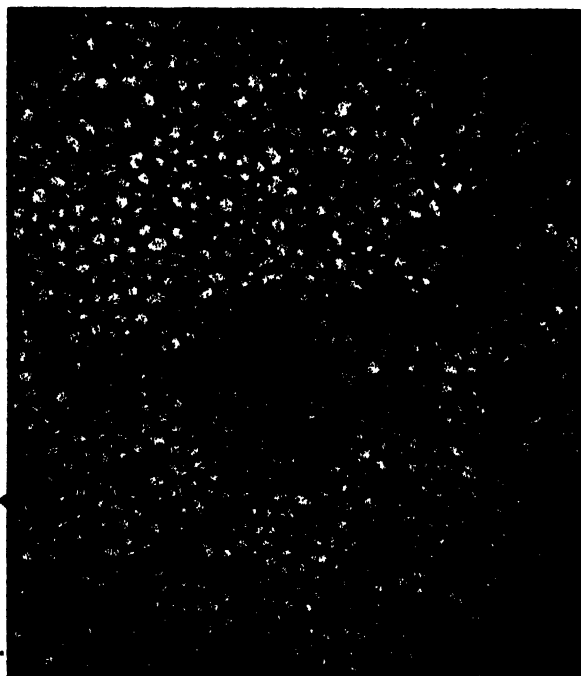


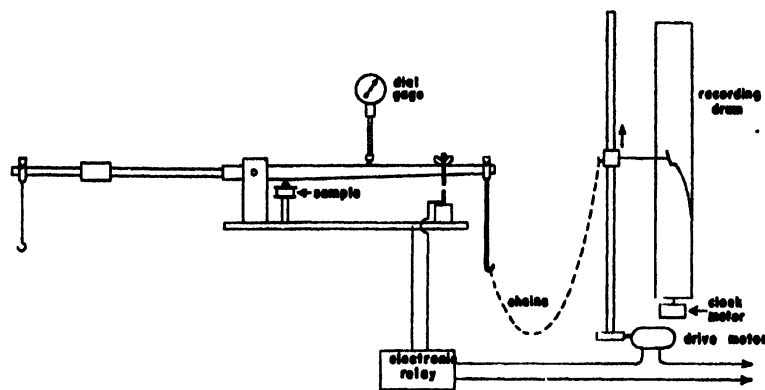
FIG. 1. Natural cork with cotton fibers. 80X.



FIG. 2. Natural cork. 800X.

* Taken from the very interesting book *Cork and the American Cork Industry* by Arthur L. Faubel, published by the Cork Institute of America in 1941.

FIG. 3.



Cork is the outer bark of an evergreen oak tree. The tree grows this outer bark as a protection against the hot dry winds which are common around the Mediterranean sea. The cork oak is native to the Mediterranean area only, although attempts have been made to introduce it into the United States with some success in California. The corky outer bark of this oak tree is not a living part of the tree and hence may be stripped off when it gets sufficiently thick, without injuring the tree. A tree cannot be stripped until it is about 30 years old and thereafter every 10 years or so. This yields a piece of cork an inch or two thick. The average life of the tree is about 150 years but many live 200 years or more.

If we take a piece of this cork and look at it under a microscope, we will notice that it is made up entirely of tiny cells, as shown in Fig. 1. Thus, cork might be thought of as a solidified foam. The black lines on the slide are cotton fibers included to give a size comparison. The cells are fairly uniform in size and average about one thousandth of an inch in diameter. It is interesting to note that the shape of these cells is tetrakaidecahedral. This is the shape as shown by Lord Kelvin that one would expect from minimum surface conditions.

The cellular structure of cork accounts for most of its useful properties. This structure explains immediately the low density and also the high coefficient of friction due to the many "cups" or opened cells at the surface. The fact that the cells are closed explains the low thermal conductivity, the resistance to moisture and liquid penetration, and to a certain extent the resilience and ability to absorb vibration. Cork is also decidedly inert chemically. Figure 2 shows the cork cells at a higher magnification. This indi-

cates the structure of the cell walls and shows the presence of crystalline material within the cells. With this brief outline of what cork is, we can proceed to a more detailed study of some of its properties.

DESCRIPTION OF EXPERIMENTAL METHOD

The apparatus used is shown schematically in Fig. 3. It consists essentially of a lever arm with a 1 to 10 ratio between the force due to the weights and that on the sample. The compression was indicated by a dial gauge on a 5 to 1 ratio with the sample. Stress-strain curves were taken by adding weights at a constant rate and reading the corresponding compression on the dial gauge. For relaxation the machine was made automatic by using the principle of the chainomatic balance. A number of chains were used as a variable weight. The moving end of the chains was driven by a motor actuated by contact points on the lever arm as indicated in Fig. 3. This set-up kept the forces balanced at all times and hence gave a very accurate picture of the relaxation. The motion of the chain drive was recorded by a pen on a clock driven drum. Thus, the machine recorded a continuous force *vs.* time curve.

STRESS-STRAIN

Because of the cellular structure of cork one can explain qualitatively the shape of the stress-strain curve in compression. Figure 4 shows the stress-strain curve for cork. At low compressions there is a region in which Hooke's law holds. At stresses slightly higher there seems to be a general breakdown of the structure and hence a region where compression changes rapidly, with small stress changes. Beyond this is the final region where the cells have all been collapsed and the material itself is compressed. This, of course,

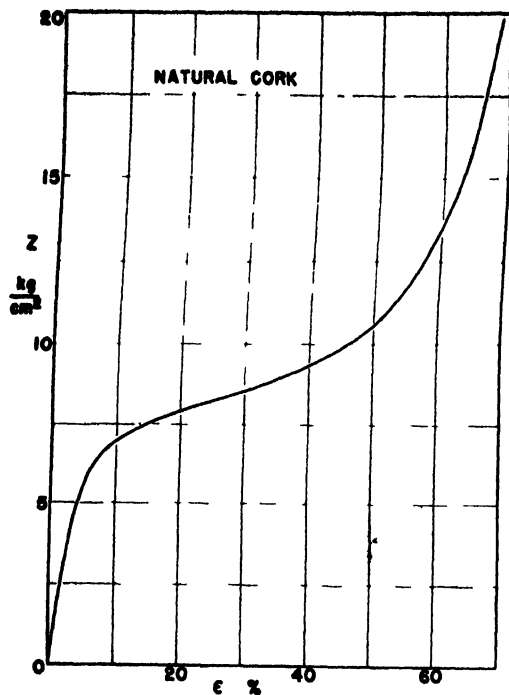


FIG. 4.

results in a final steep upward trend in the curve. Cork is not the only substance whose stress-strain curve is of this shape. Most woods, when compressed perpendicular to the grain (and hence in a direction to collapse the cells), give a similar curve. Dry Balsa wood in particular gives a curve with a very abrupt breakdown of the cells and hence that portion of the curve is almost horizontal. Metals also show an S-shaped compression curve, but here the final upward trend is due to a change in cross section of the metal. However, the unique thing about cork is that after a deformation it recovers its original dimensions to a remarkable degree. This is what gives cork its resilience and makes it useful for gaskets, stoppers, etc. We shall mention the stress-strain curve again a little later.

RELAXATION

The stress vs. time curves for natural cork in an ordinary plot have much the same appearance as that of many materials, i.e., a decay of stress whose rate of decay decreases with increasing time. Many types of plots were tried and it was found that if we plot the stress against the logarithm of the time we get almost exactly a straight line. Such a plot for 2 percent compression is shown in Fig. 5. This behavior was

checked for times of as long as a couple of weeks without deviations from linearity. At high compressions it is found that there are deviations for short times (less than 10 minutes) which may be due to the air diffusing out of the compressed cork cells. In these cases we have taken the linear behavior of the sample after the first 10 minutes as the true cork behavior.

Figure 6 shows a number of relaxation curves at various degrees of compression plotted on a semi-log graph. One can immediately see a number of interesting things. The intercept values at t equal to one minute show a variation with compression entirely similar to that shown on the stress-strain curve; in fact, within the experimental error (due to the use of many different samples) these are identical with the stress-strain curve. Although one would expect this behavior, it is not an obvious conclusion since the stress-strain curve previously shown was obtained by an entirely different process and used but one sample. Another interesting thing about this graph is the fact that the extrapolation

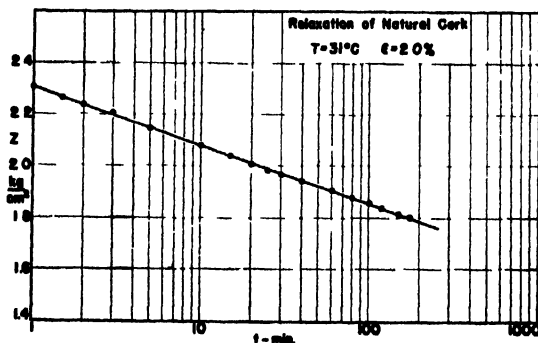


FIG. 5.

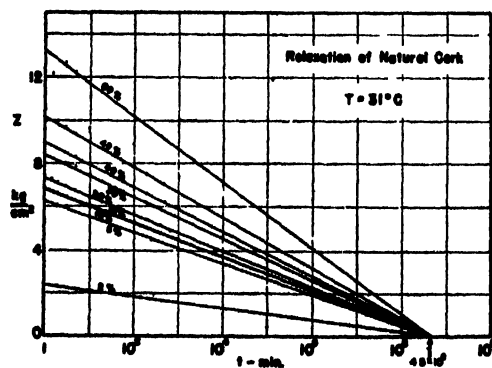


FIG. 6.

of the straight lines to zero stress yields the same decay time τ for all curves. Thus, assuming the extrapolation to be correct, we can say that no matter what the compression, the stress will decay to zero in about 980 years.

Let us see what conclusions we can draw from this fact. The equation of a single relaxation curve must be of the form.

$$Z = A_{(\epsilon, T)} - B_{(\epsilon, T)} \log_{10} t, \quad (1)$$

where Z is the stress, t is the time in minutes, and $A_{(\epsilon, T)}$ and $B_{(\epsilon, T)}$ are functions of the compression and temperature. We have already seen that the dependence of $A_{(\epsilon, T)}$ the intercept, on the compression, ϵ , is given by the stress-strain curve. Now, since all curves pass through the same point and since this point is at $Z=0$, we can say that the slope of the lines must be proportional to the intercept for all compressions. To state this a little differently, we can say that the ratio of $B_{(\epsilon, T)}$ to $A_{(\epsilon, T)}$ must be independent of compression. Thus, we may say that

$$B_{(\epsilon, T)} / A_{(\epsilon, T)} = \lambda(T), \quad (2)$$

where $\lambda(T)$ is a function of the temperature only. Rewriting Eq. (1) we have

$$Z = A_{(\epsilon, T)}(1 - \lambda(T) \log t) = F(\epsilon, T)G(t, T). \quad (3)$$

Thus, ignoring the temperature dependencies for the moment, we have split the stress function into two factors, the first depending on the compression only and the second depending on the time only. The form of the first factor, i.e., the stress-strain curve was shown graphically and explained qualitatively. The second factor was explained quantitatively and is given in analytical

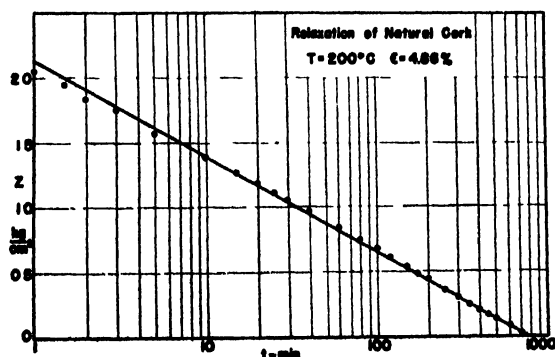


FIG. 7.

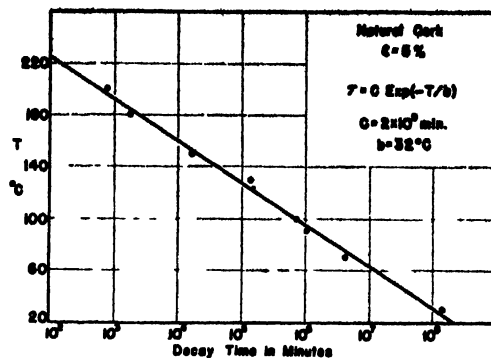


FIG. 8.

form. It should be pointed out that the decay time τ is related to λ by the equation

$$\tau = 10^{1/\lambda}.$$

TEMPERATURE DEPENDENCY

There remains to be found the temperature dependencies. One would expect that a process like relaxation would be accelerated by an increase in temperature. This proves to be the case and what is more interesting is the fact that the form is unchanged. That is, the relaxation curves are found to be linear on semi-log plot for all temperatures studied. Figure 7 shows the relaxation curve taken at 200°C. The curve is seen to be practically linear all the way to zero stress. Since raising the temperature means increasing the rate of relaxation, it seems that it was a good assumption that the semi-log plot is linear all the way to zero stress.

Figure 8 shows a plot of the temperature *vs.* the log of the decay time. There is a spread of the experimental points but this seems to be a straight line. The equation of this line would be of the form

$$\tau = \exp_{10} \left(\frac{\alpha - T}{\beta} \right) = C e^{-T/\beta}. \quad (4)$$

A theoretical consideration would predict a linear relation between $\log \tau$ and $1/T$. However, this does not check the experimental evidence as well as the above form.

Thus, we have as the generalized equation of state of cork

$$Z_{(\epsilon, T, t)} = A_{(\epsilon, T)} \left(1 - \frac{\beta}{\alpha - T} \log t \right). \quad (5)$$

The analysis is not quite complete as we have not given the stress temperature relationships. There

has been a great deal of experimental difficulty in determining this. Besides the usual difficulties of variations between samples and the large number of variables encountered (i.e., density and humidity as well as stress, temperature, time, and compression) we have to correct for thermal expansion *and rate of thermal expansion* in the apparatus. However, the experimental difficulties have been pretty well cleared up and it is hoped that this last phase of the study will be completed soon.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. H. A. Robinson, Chief Physicist, Armstrong Cork Company for valuable discussions and for furnishing the samples used in this investigation. They also appreciate the benefit of discussions and the support of this work by Mr. E. Claxton, Director of Research and Dr. R. B. Rohrer, Assistant Director of Research, Armstrong Cork Company.

The Formation of Ionized Water Films on Dielectrics under Conditions of High Humidity

ROBERT F. FIELD

Gen. Rad. Radio Company, Cambridge, Massachusetts

When a dielectric is placed in an atmosphere of 100 percent relative humidity, an ionized film of water forms, whose conductivity at the end of one minute is within a factor of ten of its equilibrium value, which is usually attained within an hour. This equilibrium conductivity ranges from essentially zero for certain hydrocarbon waxes, silicone resins and silicone-treated glass to 100 micromhos for ordinary glass and quartz. The ionized water film also produces interfacial polarization at its interface with the dielectric, which produces a marked increase in both capacitance and dissipation factor at audiofrequencies. This polarization builds up in the same manner as the conductivity. Its relaxation frequency appears to be in the audio-range.

THE effect of high humidity on dielectrics is so important that methods of conditioning samples have already been established. The properties studied have been mechanical, such as

tensile strength and impact strength. Hence the conditioning has been for volume absorption, not for surface film. These conditioning procedures have been carried over to the testing of

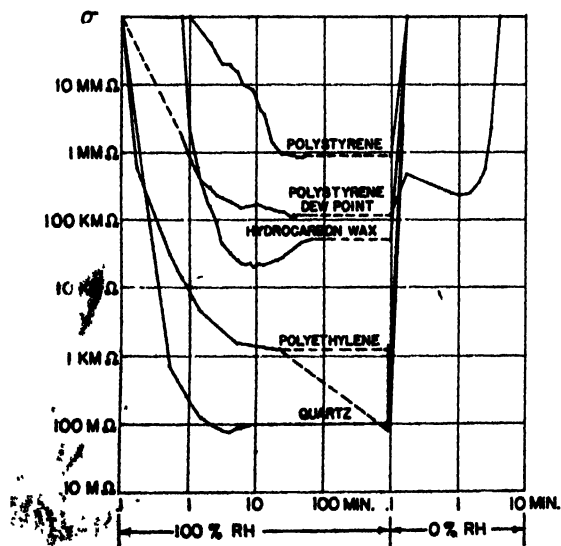


FIG. 1. Change of surface resistivity with time for materials having negligible volume absorption; exposure to 100 percent RH on left, recovery at 0 percent RH on right.

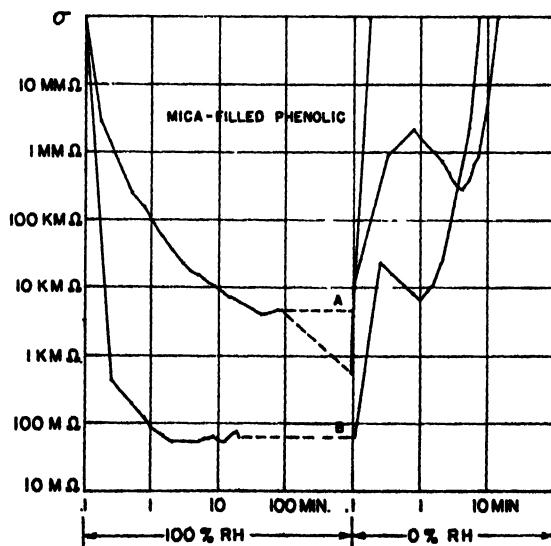


FIG. 2. Exposure and recovery curves for materials whose surface films are so thick that natural diffusion cannot cause them to vanish quickly.

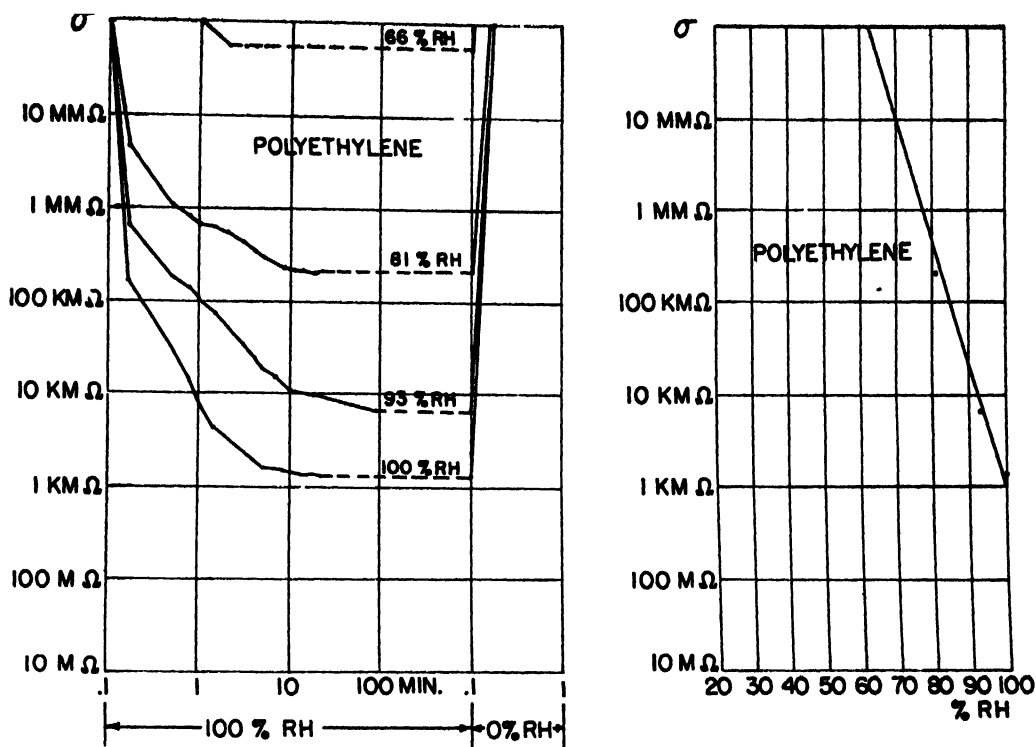


FIG. 3. (left) Effect of relative humidity on magnitude of exposure curve. (right) Relation of equilibrium surface resistivity to relative humidity.

electrical properties, where actually the effect of a surface film is of the greater immediate importance. The effect of volume absorption appears only after hours or days, while the surface film forms in minutes.

D.C. PHENOMENA

The formation of surface films has been studied by measuring the resistance between two pairs of electrodes bolted to opposite ends of thin slabs of dielectric.¹ This resistance is greater than 1 M Ω and is measured on a megohm bridge.² The slab is mounted on polystyrene terminals set in the metal top of a glass desiccator jar. Leakage over the surface of these terminals is eliminated by connecting the metal top to the guard terminal of the bridge. The base of the jar is filled with water for 100 percent RH, with saturated solutions of various salts for lower humidities³ and with silica gel for 0 percent RH.

¹ ASTM Test Method, D257-38, Figs. 1 and 2.

² General Radio Type 544-BS8 Megohm Bridge, having upper resistance limit of 10 MM Ω .

%RH 93 81 66 52 43 32
Salt Na₂SO₄ (NH₄)₂SO₄ NaNO₃ Na₂Cr₂O₇ K₂CO₃ CaCl₂

The rate of formation of surface film is shown in Fig. 1 for several dielectrics which have no appreciable volume absorption. Within a minute after being placed in 100 percent RH the surface resistivity is reduced from greater than 20 MM Ω to a value about a decade above the final equilibrium value, which is usually attained within 1 hour. Evidence to be given later suggests that the film forms very rapidly and that it is the ionization of the film that is observed by measuring resistance. Surface resistivity may decrease steadily to its equilibrium value, as for polyethylene, or overshoot, as for quartz and wax. In the latter case, excess ions are swept to the electrodes by the steady voltage. This action is well shown by polyethylene which has been allowed to stand overnight with the voltage off. Resistivity dropped by a decade, but recovered in a few minutes.

The film is actually very thin and vanishes within 10 seconds at 0 percent RH except for the thicker films which result from dew point condensation, shown here for polystyrene. In this case natural diffusion in the desiccator jar cannot

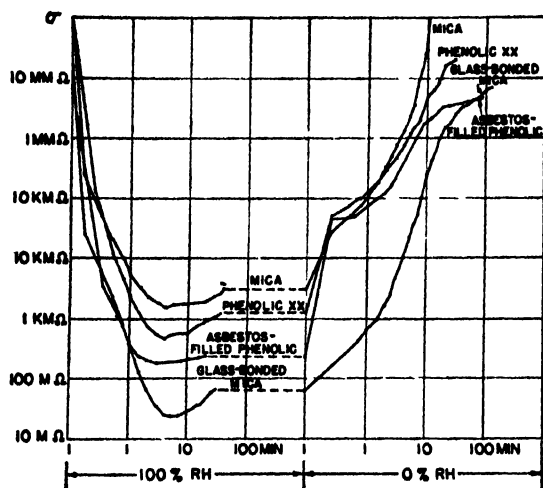


FIG. 4. Effect of volume absorption on shape of recovery curve.

maintain 0 percent RH at the surface. A curve with a maximum and minimum results, which appears to be typical. Forced ventilation removes the film within 10 seconds.

Similar typical recovery curves for two mica-filled phenolics are shown in Fig. 2. They illustrate the importance of small differences in composition or surface finish on the thickness and tenacity of the film. Type A requires 16 hours to give the typical recovery curve which Type B produces in 30 minutes. This is not caused by volume absorption, for forced ventilation removes the films from both materials in less than 10 seconds.

The effect of lower values of relative humidity is shown in Fig. 3 for polyethylene. The equilibrium value is a logarithmic function of relative humidity. This relation has been found for all the materials studied and agrees with the conclusions to be drawn from the classic work of Curtis.⁴

The effect of volume absorption first appears in the shape of the recovery curve, as shown in Fig. 4. Water stored just below the surface maintains the surface film without adding appreciable conductivity of its own. Surface forces acting on the water molecules must play a very important part, for the volume absorption of asbestos-filled or laminated phenolics differs greatly from that of natural mica and glass-bonded mica.

⁴H. L. Curtis, Bull. Bur. Stand. 2, 359 (1915).

The effect of volume absorption on the exposure curve is shown in Fig. 5 for polyamide, which has a high volume absorption as compared to its surface film. Conductivity from the absorbed water becomes comparable with that from the surface film before an equilibrium value is attained. Vulcanized fibre has so much surface film that its equally great volume absorption does not prevent equilibrium at 100 percent RH. At 81 percent RH however its surface film is reduced while the volume absorption is unchanged, so that there is no equilibrium value.

Some 40 different materials have been studied in this manner. Their parameters are shown in Table I, arranged in order of decreasing surface resistivity. For a material to be water repellant there must be some hydrophobic radical strategically placed on the surface. Methyl, ethyl, butyl, and propyl radicals seem to meet this requirement. The position of cellulose acetate butyrate as second in the list is the more remarkable because it maintains its high resistivity even after absorbing 4 percent of water. Silicone rubber and other silicone materials contain a methyl radical. Non-absorbent materials containing silicon, like quartz, glass, and steatite, can be made completely water repellant by coating with special silicone resins which provide a layer several hundred molecules thick. The exact position in the table of many of the other materials is determined by the condition of their surface. The samples measured were cleaned with grain alcohol and then baked at 60°C.

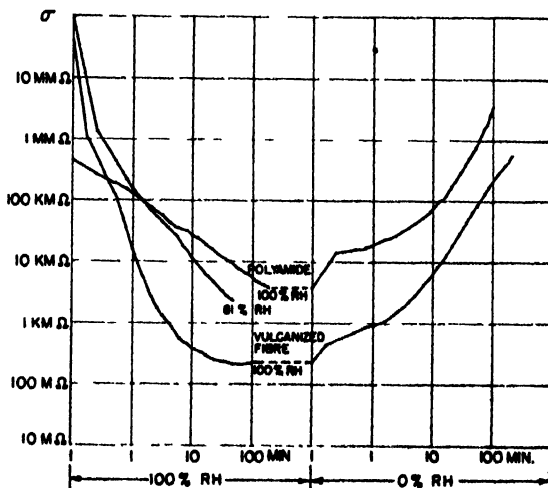


FIG. 5. Effect of volume absorption on exposure curve.

This provides as clean a surface as will exist in normal use, but does not completely remove oil films and other substances. Only one sample of a kind was tested so that there is no measure of the magnitude of normal variations. Values on the same specimen are repeatable within 50 percent.

The second column in the table gives the slope of the straight line plot of surface resistivity against relative humidity with values ranging

TABLE I.

Material	σ	% RH per decade change of σ	Recovery time in minutes after 100% RH for	
			1 hr.	16 hr.
	MΩ			
Hydrocarbon wax, modified	>20		.0	.0
Cellulose acetate butyrate	>20		.0	.0
Silicone rubber	10		.13	
Polytetrafluoroethylene	3.6		.17	.17
	kMΩ			
Polystyrene (sheet)	840		.13	.13
Polydichlorostyrene 2 5	29	7	.17	
Hydrocarbon wax	20	13	.17	
Ethyl cellulose	13	9	.33	.5
Cellulose acetate	7.0	6	1.0	6
Polyvinyl chloride acetate	5.7	12	6*	
Polystyrene (plasticized)	5.0	4	.17	62*
Phenolic, mica-filled	5.0	9	.17	13*
Aniline formaldehyde	4.2	4	.17	20*
Polyamide	3.8	14	200	
Porcelain, glazed	3.7	15	2.5*	
Glass (high K)	3.4	10	17*	20*
Mica	3.0	12	11	
Polystyrene (molded)	2.4	10	.17	.17
Polystyrene (plasticized)	2.4	8	.11	.17
Steatite (L-3)	1.6		.17	.75
Quartz	1.4			
Polyethylene	1.3	9	.17	.17
Phenolic, XX	1.3	16	80	
Phenolic, asbestos filled	1.2	9	1.5*	100*
	MΩ			
Phenolic, XXXP	660	15	300	
Steatite (L-4)	640		.5	1
Phenolic, LE	500	18	400	
Phenolic, mica-filled	320	8	40*	
Steatite (L-4)	280		.33	
Polydichlorostyrene 3-4	240	6	.33	5.3*
Phenolic, cellulose filled	240	10	400	
Aniline formaldehyde, glass				
matte	240	9	14	1000
Phenolic, C	220	16	300	
Vulcanized fibre	220		6000	
Aniline formaldehyde, glass				
cloth	200	12	3	
Quartz	190			
Phenol formaldehyde (plasticized)	100	12	25	
Glass (sintered)	90			
Glass bonded mica	64	18	400	
Melamine, glass cloth	38	14	300	
Phenolic, mica filled	30	11	7*	

Starred (*) times indicate that forced ventilation will give a recovery time of less than 0.25 minute.

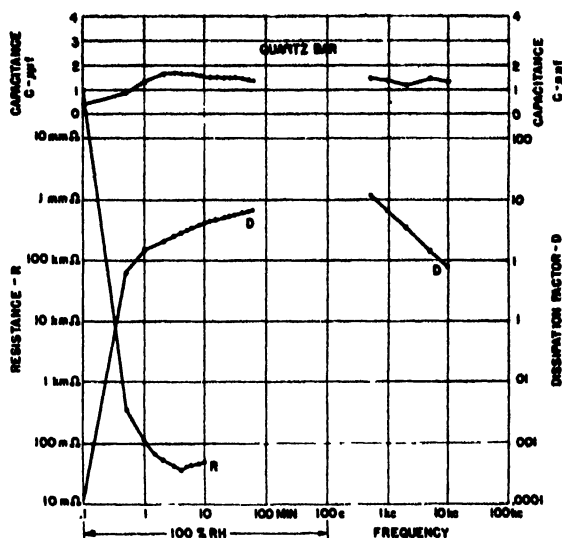


FIG. 6. Change of capacitance and dissipation factor with time and, after equilibrium, with frequency.

from 4 to 18 percent RH per decade change in resistivity. The third and fourth columns give the recovery times in minutes after exposure to 100 percent RH for 1 and 16 hours. Any time up to 0.25 minute (15 seconds) indicates no volume absorption. Any longer time which is starred (*) is for a recovery curve typical of non-porous materials, for which forced ventilation will give a recovery time of less than 0.25 minute. Other large times indicate volume absorption under the conditions noted.

A.C. PHENOMENA

The effect of surface films on the dielectric constant and dissipation factor of dielectrics has been studied by measurements on a Schering bridge,⁸ both at a single frequency as the film is forming and vanishing and over a considerable frequency range under equilibrium conditions. It is obvious that, as the film forms, both capacitance and dissipation factor should increase, because the water film itself must have a dielectric constant of 80 and must provide a shunt resistance because it is conducting.

The behavior of a quartz bar at 100 percent RH is shown in Fig. 6. Both capacitance and dissipation factor at 1 kc increase to steady values as the surface film builds up and its

⁸ General Radio Type 716-C capacitance bridge.

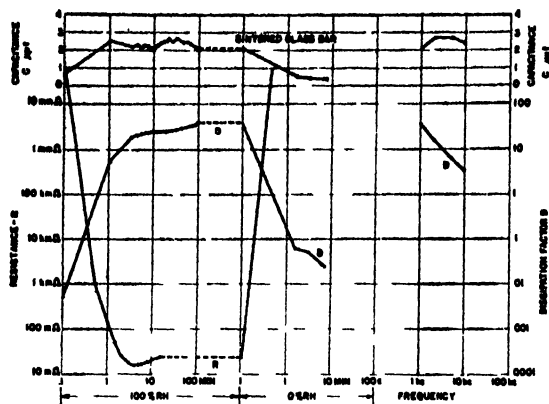


FIG. 7. Exposure, equilibrium, and recovery curves. Capacitance and dissipation factor recover more slowly than surface resistivity.

resistance decreases to its equilibrium value. For frequencies in the audio-range capacitance is con-

stant within the limits of error for such small values. Dissipation factor varies inversely with frequency and has the value to be determined by the equilibrium resistance. There is no evidence of any interfacial polarization.

Similar curves for a sintered glass bar are shown in Fig. 7. The 1 kc recovery curves, taken after equilibrium conditions are attained, show that, after the film breaks within 0.25 minute and raises surface resistance above 10 MMΩ, there are ionized water droplets on the surface, for capacitance and dissipation factor do not return to their original values for nearly 10 minutes.

The behavior of an L-4 grade steatite is shown in Fig. 8. Both capacitance and dissipation factor at 1 kc continue to rise up to the time of surface resistance equilibrium. Their equilibrium values

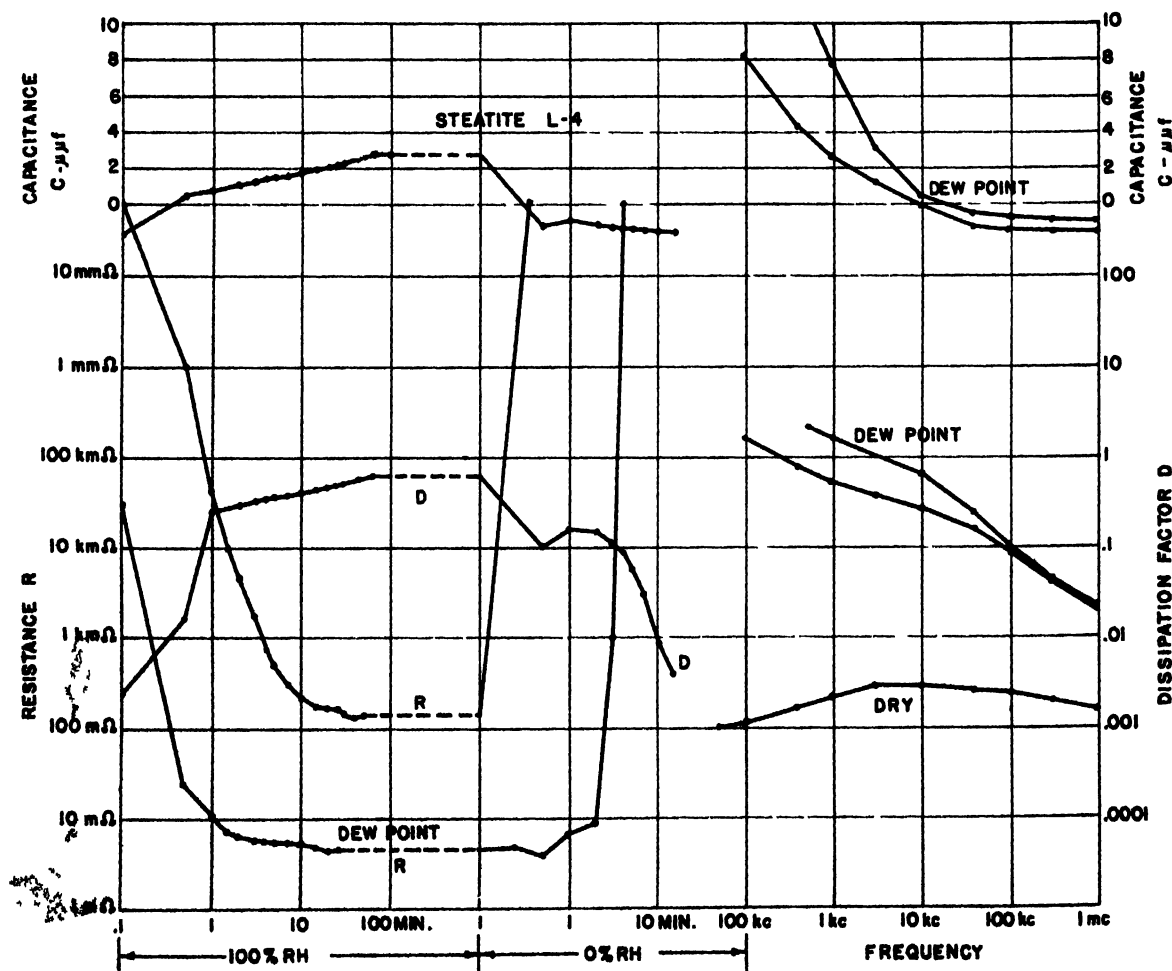


FIG. 8. Exposure, equilibrium, and recovery curves for material showing polarization at the interface of water film and dielectric. Also effect of dew-point condensation.

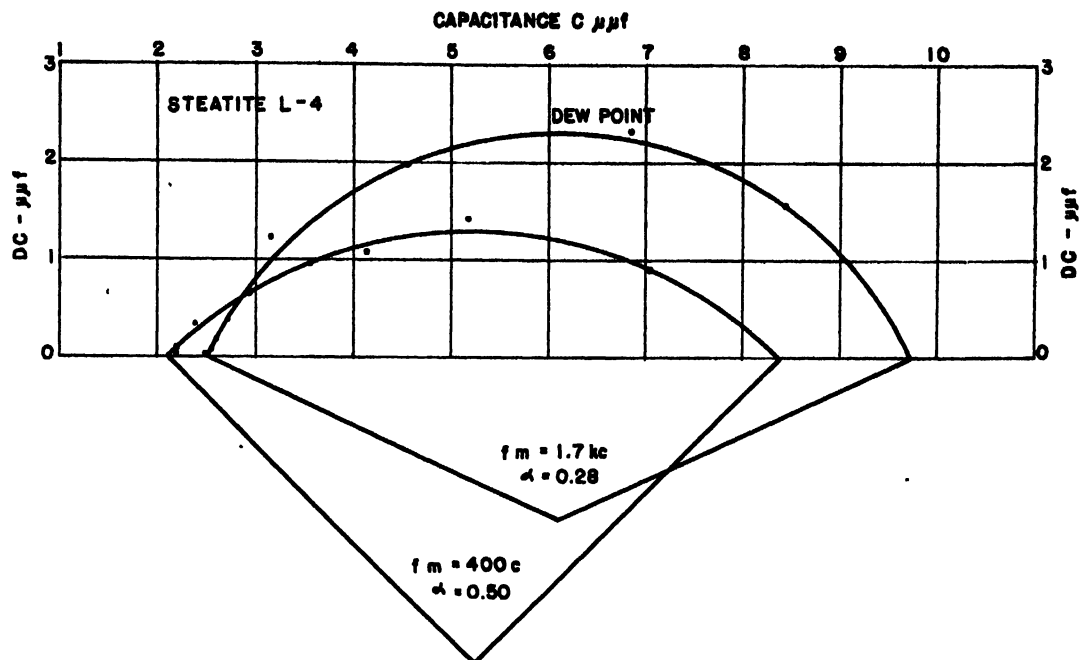


FIG. 9. Circular arcs for equilibrium data of Fig. 8.

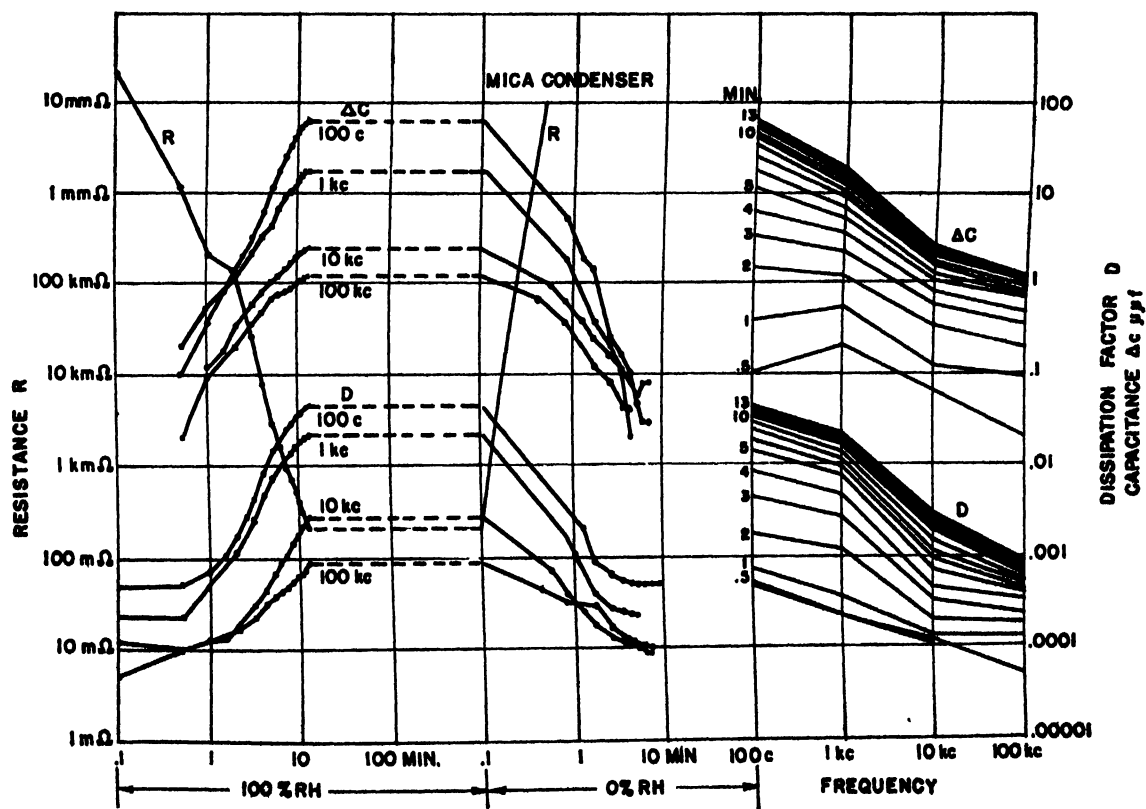


FIG. 10. (left) Exposure and recovery curves at four different frequencies plotted against time. (right) Exposure curves at one-minute intervals up to 13 minutes plotted against frequency.

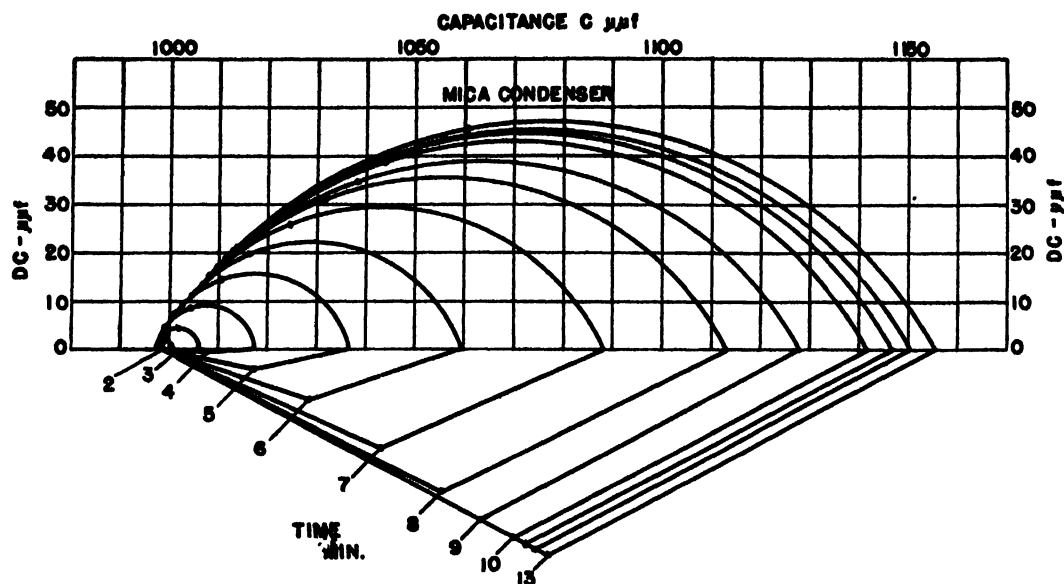


FIG. 11. Circular arcs for exposure curves at one-minute intervals.

at other frequencies definitely indicate interfacial polarization. Capacitance is no longer independent of frequency, but increases rapidly with decreasing frequency. Dissipation factor no longer decreases inversely with frequency, but changes more slowly. After subtracting the values of dissipation factor resulting from the equilibrium surface resistance, the circular arc plot⁶ of the product DC of dissipation factor and capacitance against capacitance appears as shown in Fig. 9. This is an audiofrequency polarization with a relaxation frequency f_m of 400 c, a storage coefficient α of 0.50, and a change in capacitance ΔC of $6.18 \mu\text{f}$. Observations were also taken on

the thicker film produced by dew-point condensation. Its circular arc plot as shown in Fig. 9 gives a higher relaxation frequency, 1.7 kc, a smaller storage coefficient, 0.28, and a slightly larger change in capacitance, $7.22 \mu\text{f}$. This indicates that the ion concentration is less in the thicker film, which is reasonable.

A more complete study of this type of interfacial polarization was made on an unprotected $1000 \mu\text{f}$ mica condenser.⁷ This condenser was exposed to 100 percent RH five times to obtain the resistance curve and the capacitance and dissipation factor curves at four different frequencies, as shown in Fig. 10. Only capacitance changes are given. The remarkable consistency of these successive exposures is shown by the fact that in the frequency plots at one-minute intervals no curves cross except those for the shortest times. After subtracting the effect of surface resistance the circular arc plots of Fig. 11 were obtained. There is an increase with time in all three polarization parameters, change in capacitance, relaxation frequency, and storage coefficient, the latter increasing from 0.0 to 0.31. This is the evidence which indicates that the film forms rapidly and then ion concentration slowly increases, for storage coefficient α can be zero (semicircle) only for dilute solutions. The

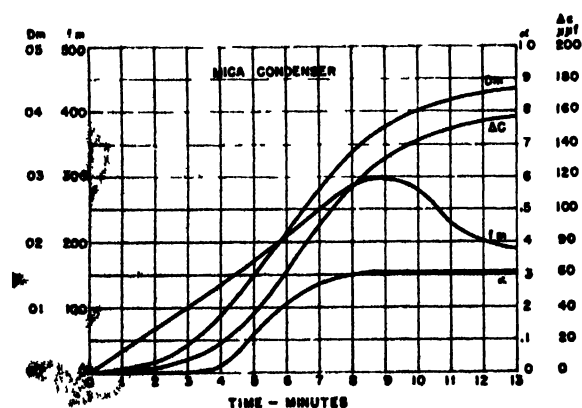


FIG. 12. Change of polarization parameters with time.

⁶ K. S. Cole and R. H. Cole, J. Chem. Phys. 9, 341 (1941).

⁷ General Radio Type 505-F, uncaused.

changes with time in the numerical values of the three parameters are shown in Fig. 12.

These a.c. measurements are far from complete, for they must be taken for most of the materials studied for d.c. leakage before broad conclusions can be drawn. It will be important to establish the order in which these materials are arranged in terms of either fractional change in capacitance or maximum dissipation factor and to determine the corresponding values of

storage coefficient and relaxation frequency in order to set a frequency limit beyond which humidity has no appreciable effect.

ACKNOWLEDGMENT

The author is indebted to Mr. D. H. Chute for mounting the various materials and for carrying out most of the measurements for both resistance and capacitance and dissipation factor.

Thermal Polymerization of Drying Oils

H. E. ADAMS AND P. O. POWERS*

Armstrong Cork Company, Lancaster, Pennsylvania

A statistical analysis of the thermal polymerization of drying oils has been made, and equations have been derived which predict the amount of a specific polymer present as a function of the extent of reaction and the original composition of the oil. Several samples of oils have been heat bodied to test the validity of this analysis. Certain discrepancies are apparent. The presence of an intrapolymer is postulated to explain these discrepancies. Indirect evidence for the presence of such a compound is presented. Additional studies of the heat polymerization of drying oils have been made by conducting the polymerization in dilute solution. After the reaction has been carried out, the solvent is removed and the properties of the oils compared with oils similarly treated without solvent. It has been found that iodine values of the oils are substantially the same regardless of the concentration of oil, and depend only on the time and temperature of heating, but that the viscosity decreases with lower concentration of the oils. This behavior substantiates the formation of an intrapolymer. By using the values at various concentrations it may be possible to estimate the extent of intrapolymer formation in heat bodied oils.

IT is well known that drying oils such as linseed, tung, and soybean oils increase in body or viscosity when heated at elevated temperatures. This increase in viscosity is due to a chemical reaction between the unsaturated fatty acids present in the oils as glycerides. Various types and varying amounts of fatty acids account for the differences in reactivity and properties of the different oils. The vegetable oils are composed chiefly of glycerides of eighteen carbon atom fatty acids, while the fish oils contain, in addition to eighteen carbon atom fatty acids, fatty acids with twenty to twenty-six carbon atoms. These fatty acids contain from none to six double bonds and in some cases an hydroxyl group is attached to the chain. In most cases the double bonds are

isolated, but conjugated double bonds do occur such as in eleostearic acid in tung oil. It has been observed that non-conjugated double bonds undergo a slow but progressive shift to the conjugated position at elevated temperatures.¹ Thus, a conjugated double bond system is postulated as a prerequisite for polymerization. Extensive work has been done on the polymerization of conjugated systems such as butadiene, isoprene, etc.² Vinylcyclohexene is obtained in the dimerization of butadiene by the addition of one mole-

¹ T. F. Bradley and W. B. Johnston, *Ind. Eng. Chem.* **32**, 802 (1940); **33**, 86-89 (1941). T. F. Bradley and D. Richardson, *ibid.* **32**, 963 (1940). J. S. Brod, W. G. France, and W. L. Evans, *ibid.* **31**, 114 (1939).

² F. Hofmann and L. Tank, *Zeits. f. angew. Chemie* **25**, 1465 (1912). S. Lebedev and S. Sergienko, *Comptes rendus Acad. Sci. U.S.S.R.* **3**, 79 (1935). K. Alder and H. F. Rickert, *Ber. d. d. Chem. Ges.* **71B**, 373 (1938). G. S. Whitby and W. Gallay, *Can. J. Research* **6**, 280 (1932).

* Present Address: Battelle Memorial Institute, Columbus, Ohio.

cule of butadiene in the 1-4 position to another molecule in the 1-2 position. This Diels-Alder reaction is taken as the mechanism of the polymerization that takes place during heat bodying of drying oils. It is possible that three or more fatty acids could combine. The formation of trimer fatty acids has been suggested,³ but the extent of their formation in heat bodying has not been established. Since it is not the free fatty acid but the glyceride of the fatty acids which occurs in the drying oils, highly branched structures or molecules are formed during the heat bodying process.

COMPOSITION OF LINSEED OIL

It is fairly well established⁴ that linseed oil consists of glycerides of eighteen carbon atom acids and that the chief acids present are stearic, oleic, linoleic, and linolenic. Since the iodine value of linseed oil varies somewhat, the relative amounts of the various fatty acids must vary accordingly. Equations⁴ have been proposed to correlate the iodine number with the composition of the oil. Two recent articles have studied the percentages of the various fatty acids in linseed oil by the determination of thiocyanate and iodine values of the oils and by determining the ultra-violet absorption of the alkali isomerized acids. The work of Mitchell, Kraybill, and Zscheile⁵ has shown that estimates of the linolenic and linoleic acid content, by determining absorption at 2340 and 2680A after isomerization with alkali at 180°C in glycol, give good agreement with the Kaufman method, which depends on determination of the thiocyanogen value. Their estimate for a linseed oil having a Wijs iodine value of 180 is:

Linolenic acid	47 percent,
Linoleic	18 percent,
Oleic acid	29 percent,
Saturated acids	6 percent.

These results indicate a much higher content of oleic acid and a much lower content of linoleic

acid than the prior estimates. This difference is rather large and would suggest a much lower degree of reactivity. Using the thiocyanogen value, Painter and Nesbitt⁶ have examined samples of linseed oil varying in iodine value from 127 to 203. Saturated acids were determined by the Bertram method. They did not attempt to correlate the composition of the oils with their iodine values.

It is obvious that the saturated fatty acids cannot polymerize, and since oleic acid does not readily undergo polymerization, the acid distribution may be simplified by considering two types of acids, those that undergo polymerization and those that do not. Thus, oleic acid and the saturates would be considered as non-polymerizing acids (v), and linoleic acid and linolenic acid would constitute the polymerizing acids (u). The assumption that the linoleic acid group has the same reactivity as the linolenic acid group is highly questionable, but for the present it will be assumed to be true. This point will be discussed further, later in the article. According to the analysis of Mitchell, referred to above, linseed oil fatty acids would be comprised of 65 percent polymerizable acids and 35 percent non-polymerizable acids. If the fatty acids can be considered to be randomly distributed⁷ in the glycerides, the relative amounts of the different types of glycerides are given by the following equation.

$$1 = (u+v)^3 = u^3 + 3u^2v + 3uv^2 + v^3. \quad (1)$$

Thus, (u^3) is the fraction of glyceride molecules containing three polymerizing acids, ($3u^2v$) the fraction of glyceride molecules containing two polymerizing acids and one non-polymerizing acid, ($3uv^2$) the fraction of glyceride molecules containing one polymerizing acid and two non-polymerizing acids, and (v^3) is the fraction of glyceride molecules containing three non-polymerizing acids. Thus, the drying oils according to this analysis are composed of four types of molecules, trireactive (u^3), di-reactive ($3u^2v$), mono-reactive ($3uv^2$), and zero-reactive (v^3). If the distribution is assumed to be random, the glyceride composition of some of the more im-

³ J. C. Cowan, L. B. Falkenburg, and H. M. Teeter, Ind. Chem. Anal. Ed. 16, 90 (1944). W. C. Ault, J. C. and J. P. Kass, and J. E. Jackson, Ind. Eng. Chem. 34, 1770 (1942).

⁴ P. O. Powers, Oil and Soap 22, 52 (1945). E. P. Painter, *ibid.* 21, 345 (1944). H. R. Sallons and G. D. Sinclair, Can. J. Research 22F, 132 (1944).

⁵ J. H. Mitchell, H. R. Kraybill, and F. P. Zscheile, Ind. Eng. Chem. Anal. Ed. 15, 1 (1943).

⁶ E. P. Painter and L. L. Nesbitt, Ind. Eng. Chem. Anal. Ed. 15, 123 (1943).

⁷ H. E. Longenecker, Chem. Rev. 29, 201 (1941).

TABLE I. Glyceride composition of various drying oils.

Oil	u	v	u ²	3uv	3v ²	v ³
Linseed oil	0.65	0.35	0.275	0.444	0.238	0.043
Cottonseed oil	0.43	0.57	0.082	0.323	0.416	0.180
Tung oil	0.79	0.21	0.496	0.384	0.100	0.009
Perilla oil	0.84	0.16	0.602	0.333	0.061	0.004
Oiticica oil	0.84	0.16	0.595	0.338	0.064	0.004
Soybean oil	0.61	0.39	0.227	0.436	0.278	0.059

portant drying oils are listed in Table I. The original fatty acid compositions are according to Mattil.⁸

Let the subscripts 100, 010, and 001 refer to trireactive, direactive, and monoreactive glycerides, respectively. Then the original mole fractions (N^0) and group fractions (M^0) of the different types of glyceride are given by the following equations. By group fraction is meant the fraction of reactive groups that is attached to the particular glyceride.

$$\left. \begin{aligned} N^0_{100} &= \frac{n^0_{100}}{n^0} = u^3 \\ N^0_{010} &= \frac{n^0_{010}}{n^0} = 3uv^2 \\ N^0_{001} &= \frac{n^0_{001}}{n^0} = 3uv^2 \end{aligned} \right\} \quad (2)$$

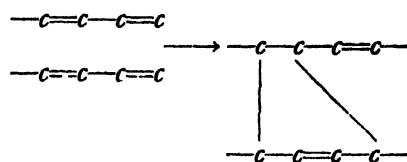
$$\left. \begin{aligned} M^0_{100} &= \frac{3n^0_{100}}{m^0} = u^2 \\ M^0_{010} &= \frac{2n^0_{010}}{m^0} = 2uv \\ M^0_{001} &= \frac{n^0_{001}}{m^0} = v^2 \end{aligned} \right\} \quad (3)$$

The functionality (f) is defined as the average number of reactive groups per molecule. That is,

$$f = \frac{m^0}{n^0} = 3N^0_{100} + 2N^0_{010} + N^0_{001} = 3u. \quad (4)$$

MECHANISM OF POLYMERIZATION

During heat bodying it is considered that these three types of glycerides combine to form polymers. It is fairly well established, although not definitely proven, that the bond between two fatty acids is a six-membered ring formed by the following mechanism.⁹



All the changes in physical properties are in accord with this mechanism, but they do not necessarily eliminate other mechanisms. Irrespective of the mechanism it is sufficient for present purposes to know that some sort of bond is formed between the reactive fatty acid groups.

The equations giving the relative amounts of the different types of polymers formed can be derived if the following assumptions are made, that during the polymerization two reactive groups disappear in the formation of one bond between two molecules, that the formation of a ring structure or intrapolymer is excluded, and finally, that all of the groups are equally reactive. The equivalence of linolenic and linoleic acid in the polymerization reaction has been previously discussed. The second assumption of the absence of ring formation is very doubtful due to the formation of highly branched structures. It is impossible to estimate the amount of ring formation. Consequently, it will be neglected for the present.

For the formation of a bond, it is necessary that two molecules collide in such a position for the reaction to take place. Thus, if we assume that the polymerization reaction is second order, the following differential equation for the instantaneous change of a definite polymer molecule is obtained.

$$\frac{dn_{abc}}{dt} = m^0 k \sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c m_{i,j,k} m_{a-i, b-j, c-k} - m^0 k m_{abc} \sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c m_{i,j,k} \quad (5)$$

⁸ W. H. Mattil, Oil and Soap 21, 197 (1944).

⁹ C. P. A. Kappelmeier, Farben Ztg. 38, 1018, 1077 (1933).

where n_{abc} is the number of molecules made from (a) trireactive glycerides, (b) direactive glycerides, and (c) monoreactive glycerides, and m_{abc} is the number of reactive groups attached to (abc) polymers. k is the specific rate constant. The first term of the right-hand side of Eq. (5) accounts for the synthesis of the (abc) polymer from all of the possible combinations of lower polymers, while the second term accounts for the disappearance of the (abc) polymer as it combines with other molecules to form higher polymers. Equation (5) may be expressed in terms of group fractions (M) and extent of reaction (p), thus eliminating the time factor, by means of the following equations

$$m = m^0(1-p), \quad (6)$$

$$dp/dt = m^0 k (1-p)^2, \quad (7)$$

$$M_{abc} = \frac{(a-c+2)n_{abc}}{m}. \quad (8)$$

Therefore,

$$\frac{d(M_{abc})}{dp} + \frac{(a-c+1)}{1-p} M_{abc} = \frac{(a-c+2)}{1-p} \sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c M_{ijk} M_{a-i, b-j, c-k}. \quad (9)$$

The following equation is a solution of this linear differential equation

$$M_{abc} = K_{abc} p^{a+b+c-1} (1-p)^{a-c+1}, \quad (10)$$

where

$$K_{abc} = \frac{a-c+2}{2(a+b+c-1)} \sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c K_{ijk} K_{a-i, b-j, c-k}, \quad (11)$$

and

$$k_{000} = 0,$$

$$k_{100} = M^0_{100},$$

$$k_{010} = M^0_{010},$$

$$k_{001} = M^0_{001}.$$

It is evident that (a) and (b) can have any value from 0 to ∞ , while (c) can only assume values from 0 to (a+2).

It has been found, by trial and error, that it is possible to solve the above recursion formula involving the constants. The following result is obtained.

$$K_{abc} = \frac{(2a+b)!}{(a-c+1)!} \frac{(M^0_{100})^a}{a!} \frac{(M^0_{010})^b}{b!} \frac{(M^0_{001})^c}{c!}. \quad (12)$$

Then the complete expression for the group fraction of an (abc) combination is given by

$$M_{abc} = \frac{(2a+b)!}{(a-c+1)!} \frac{(M^0_{100})^a}{a!} \frac{(M^0_{010})^b}{b!} \frac{(M^0_{001})^c}{c!} p^{a+b+c-1} (1-p)^{a-c+1}. \quad (13)$$

In addition the molecular weight on a number basis (A_n) and on a weight basis (A_w) are given by the following equations after the proper summations are made.

$$A_n = \frac{f}{1 - \frac{f}{2}p}, \quad (14)$$

$$A_w = A_1 \left(1 + \frac{f_p}{1 - (1 + M_{100}^0 - M_{001}^0)p} \right). \quad (15)$$

The same method of attack may be extended to cover the polymerization of any mixture of other polyfunctional molecules. The differential equation and its solution follow.

$$\frac{d(M_{x_1 \dots x_k})}{dp} + \frac{(Y_{x_1 \dots x_k} - 1)}{1 - p} M_{x_1 \dots x_k} = \frac{Y_{x_1 \dots x_k}}{1 - p} \sum_{i_1=0}^{x_1} \dots \sum_{i_k=0}^{x_k} M_{i_1 \dots i_k} M_{(x_1-i_1) \dots (x_k-i_k)}, \quad (16)$$

$$M_{x_1 \dots x_k} = \frac{(Z_{x_1 \dots x_k} - 1 + Y_{x_1 \dots x_k} - 1)!}{(Y_{x_1 \dots x_k} - 1)!} \prod_{i=1}^{x_k} \frac{(M_{0 \dots 1 \dots 0}^0)^{x_i}}{x_i!} p^{x_1 \dots x_k - 1} (1 - p)^{Y_{x_1 \dots x_k} - 1} \quad (17)$$

where

$$Z_{x_1 \dots x_k} = \sum_{i=1}^k x_i,$$

$$Y_{x_1 \dots x_k} = 2 + \sum_{i=1}^k (i-2)x_i.$$

$(Z_{x_1 \dots x_k})$ is the number of original molecules making up the $(x_1 \dots x_k)$ molecule and $(Y_{x_1 \dots x_k})$ is the number of reactive groups attached to the $(x_1 \dots x_k)$ polymer.

Likewise, the molecular weights are given by Eqs. (18) and (19)

$$A_n = \frac{A_1}{1 - \frac{f}{2}p}, \quad (18)$$

$$A_w = A_1 \left(1 + \frac{f_p}{1 - p(1 + \sum_{i=1}^k (i-2)M_{0 \dots 1 \dots 0}^0)} \right). \quad (19)$$

The fundamental Eqs. (13) and (19) are in agreement and substantiate the earlier work of Flory¹⁰ and Stockmayer.¹¹

EXPERIMENTAL PROCEDURE

In an effort to check the validity of this analysis several oils have been heat bodied. Three samples of linseed oil, one sample of linseed oil plus 1 percent litharge, one sample of soybean oil, and one sample of tung oil have been bodied

under a vacuum or inert atmosphere. Samples were taken at various times of reaction, and physical constants such as viscosity, molecular weight, density, etc., determined. The results are shown in Figs. 1 to 6 as a function of time. Table II contains the data obtained for the tung oil and soybean oil, while the results for linseed oil are given in an earlier report.¹² Reference to this report should be made for methods of determining the physical constants.

The bodying of the soybean oil was stopped after it was discovered that the viscosity had

¹⁰ P. J. Flory, J. Am. Chem. Soc. 63, 3083, 3091, and 3096 (1941).

¹¹ W. H. Stockmayer, J. Chem. Phys. 11, 45 (1943); 12, 125 (1944).

¹² H. E. Adams and P. O. Powers, Ind. Eng. Chem. 36, 1124 (1944).

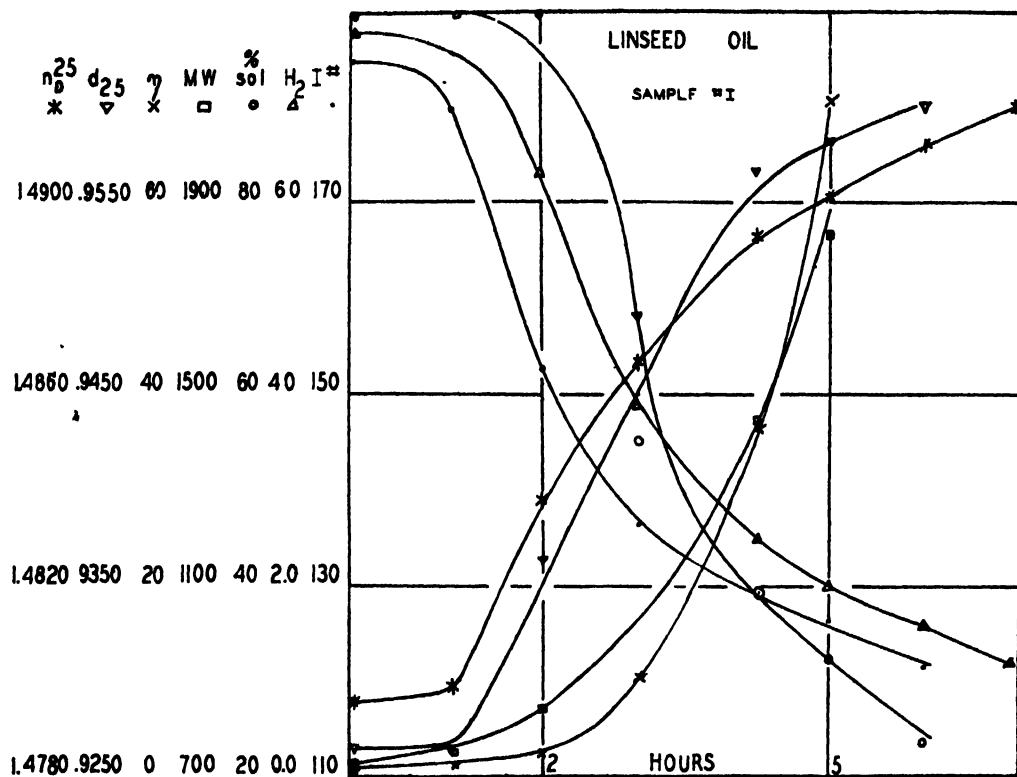


FIG. 1.

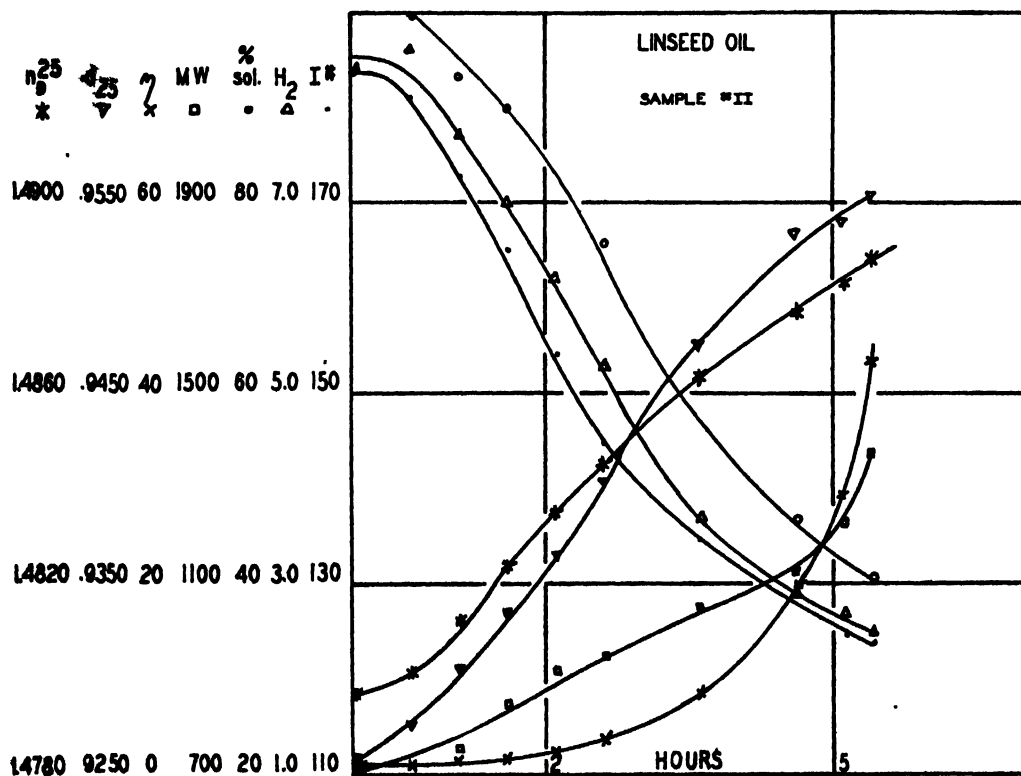


FIG. 2.

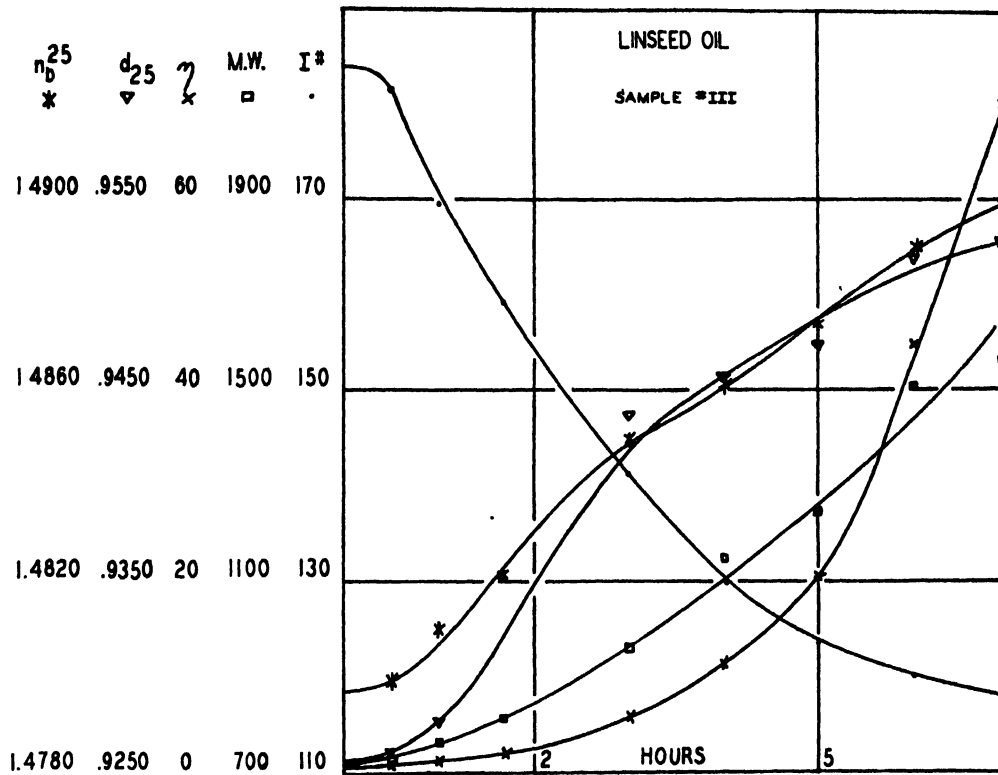


FIG. 3.

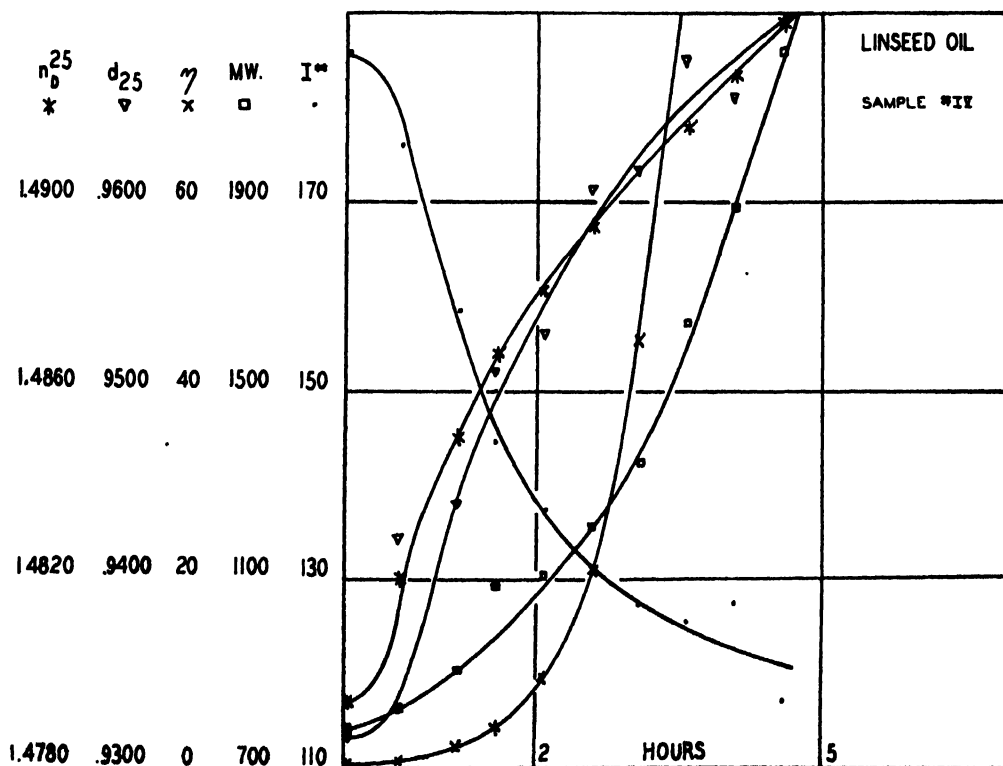


FIG. 4.

ACID %	D_{25} °	η cP	MW g	I ^o
-----------	---------------	--------------	---------	----------------

60	940	60	1300	130
----	-----	----	------	-----

40	930	40	1100	110
----	-----	----	------	-----

20	920	20	900	90
----	-----	----	-----	----

0	910	0	700	70
---	-----	---	-----	----

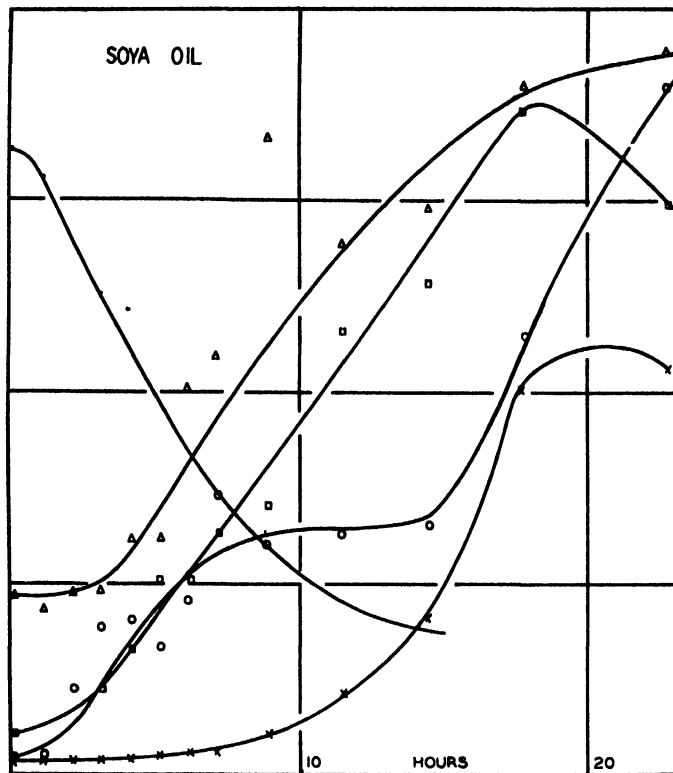


FIG 5

N_D^{25} x	H_2 °	I ^o
-----------------	------------	----------------

1534	9	170
------	---	-----

1530	7	150
------	---	-----

1526	5	130
------	---	-----

1522	3	110
------	---	-----

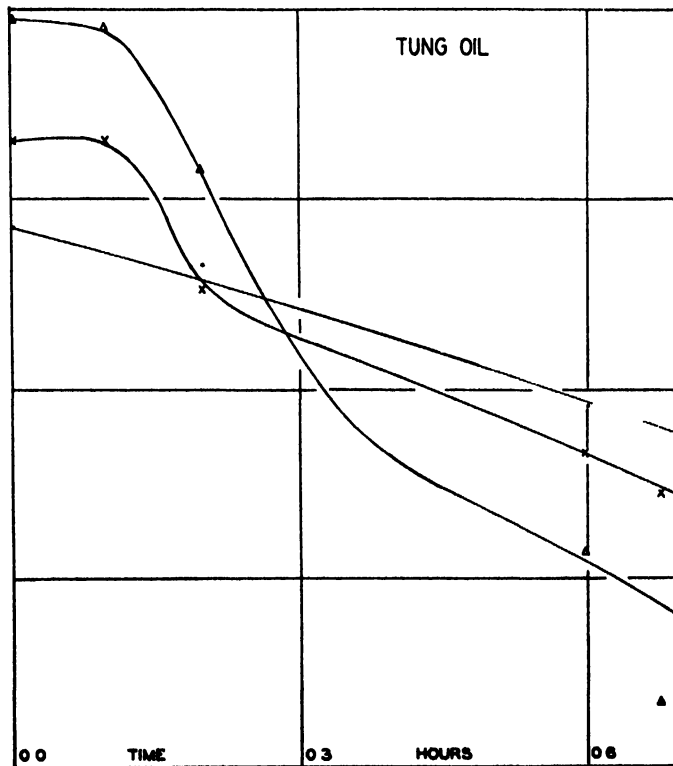


FIG 6

started to decrease and the acid number to increase greatly, showing that the glycerides were being hydrolyzed.

During the heat bodying the temperature of the reaction was controlled manually by means of a ring burner. Because of the extremely high temperature of 305°C it was difficult to maintain an entirely constant temperature. However, the maximum variation was seldom over two degrees.

DISCUSSION OF RESULTS

Rate of Reaction

Since the iodine number may be accurately determined and measures the unsaturation present in bodied oils, the values have been used to measure the extent of the combination of fatty acid groups (p). Since the loss of one double bond per fatty acid chain is equivalent to a decrease of 86.6 in iodine number, the actual decrease divided by 86.6 indicates the extent of reaction. Consequently, $(1-p)$ is a measure of the remaining double bonds or the reactivity left. Since it is postulated that in the disappearance of a double bond two fatty acid chains are bonded together through a six-membered ring, the rate of disappearance of unsaturation should vary as the

TABLE II.

Soybean oil bodied at 305°C under nitrogen						
Sample No.	Time hr.	Density	Viscosity sec.	Iodine No.	Mole weight	Acid No.
V-1	0	0.9197	1.3	135.2	746	2.0
-2	1.167	0.9188	1.3	132.4	571	2.3
-3	2.167	0.9198	1.4	120.6	572	9.4
-4	3.167	0.9198	1.7	120.3	792	15.7
-5	4.167	0.9226	1.9	117.1	834	16.5
-6	5.167	0.9226	2.1	113.0	902	13.3
-7	6.167	0.9304	2.4	99.5	903	18.3
-8	7.167	0.9320	2.6	97.9	956	29.2
-9	8.167	0.9433	4.6	95.5	984	24.5
-10	11.417	0.9379	8.7	87.6	1165	25.4
-11	14.417	0.9399	16.5	85.0	1219	26.4
-12	17.750	0.9461	40.6	—	1395	46.0
-13	22.750	0.9479	42.9	—	1299	72.0

Tung oil bodied at 225°C under vacuum				
Sample No.	Time hr.	Refractive index	Iodine No.	H ₂ absorption lb.
VI-1	0.0	1.5352	167	10.9
-2	0.1	1.5351	163	10.8
-3	0.2	1.5321	163	9.3
-4	0.6	1.5287	148	5.3
-5	0.68	1.5278	147	3.7
-6	0.70 (gel)	—	—	2.5

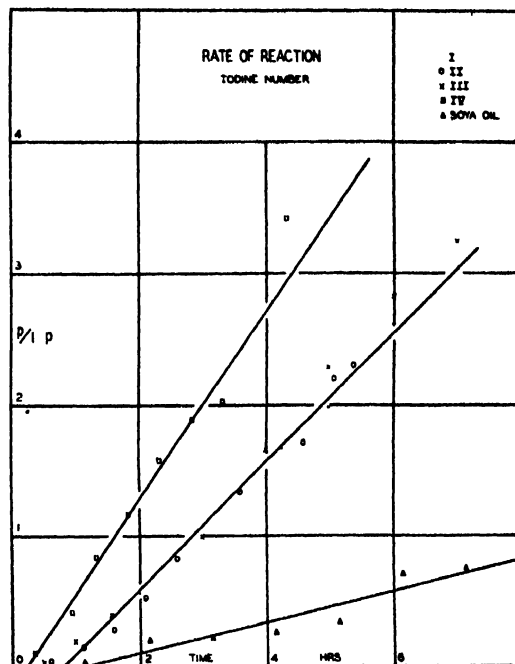


FIG. 7.

second power of the remaining reactivity. That is

$$dp/dt = k(1-p)^2, \quad (20)$$

where k is the second-order rate constant. Integration of this equation gives

$$p/(1-p) = kt. \quad (21)$$

In Fig. 7 ($p/(1-p)$) has been plotted against the time (t) for the linseed oil and soybean oil samples. Because of the high amount of conjugation present in tung oil, Wijs iodine numbers do not give a true indication of the change in unsaturation. Consequently, these results have not been included in this graph. The results clearly indicate that the disappearance of unsaturation as indicated by the iodine number is a second-order reaction. The three linseed oil samples have the same rate constant of $k=0.49$, while the presence of 1 percent litharge increases this rate constant to $k=0.73$. Soybean oil has a much smaller constant of $k=0.40$.

Mattiello¹² has bodied linseed oil at three different temperatures and determined the resulting changes in the various physical properties. With these data the extents of reaction have been

¹² J. J. Mattiello, *Protective and Decorative Coatings* (John Wiley and Sons, Inc., New York, 1943), Vol. III, p. 50.

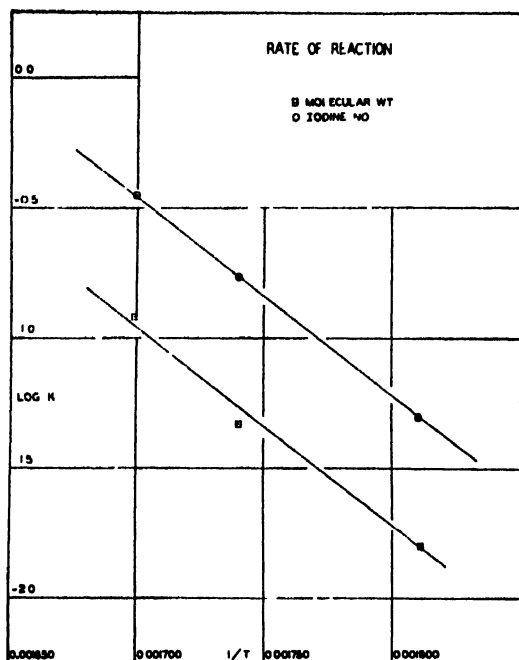


FIG. 8.

calculated from the change in iodine number and change in molecular weight. Straight lines were obtained when $(p/1-p)$ obtained from the change in iodine number and (p) as calculated from the molecular weight were plotted against time. Two parallel lines were obtained when the logarithm of the respective rate constants was plotted against the reciprocal of the absolute temperature. As shown in Fig. 8 this gave a value of 36.5 kcal. for the activation energy, as obtained from both the iodine value change and the molecular weight change. The value of 36.5 kcal. for the energy of activation is in good agreement with the estimate of the energy change (34.4 kcal.) involved in the formation of the six-membered ring as determined from the bond energies.¹⁴

Viscosity

An unsuccessful attempt has been made to obtain a relationship between the viscosity and molecular weight by means of the above equations. Flory's¹⁵ equation, which indicates that the logarithm of the viscosity is a linear function of

the square root of the molecular weight on a weight basis, does not hold for the bodied oils. A possible explanation is that the linseed oil polymer is highly branched, while Flory's equation may only be true for linear polymers.

H. Eyring¹⁶ has developed a theory of viscous flow based on reaction kinetics. Viscous flow in liquids takes place by the movement of molecules into "holes" in the liquid which in turn is governed by an energy barrier ΔE or energy of activation. This energy of activation has been found to be roughly equal to $\frac{1}{4}$ the molar heat of vaporization, consequently, it has been possible to estimate the number of molecules comprising a segment of flow. Flory¹⁵ in his work with polyesters has shown that the energy of activation is independent of molecular weight. Therefore, the segment of flow is constant. There are about 28-34 chain atoms per unit of flow.

The viscosities of heat bodied linseed oil at several stages of reaction have been measured over the temperature range of 30°C to 75°C by means of Gardner-Holt viscosity tubes. The temperature of the bath was maintained constant

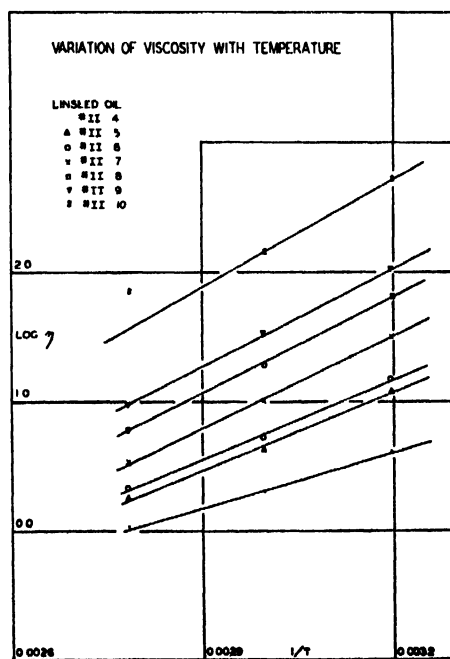


FIG. 9.

¹⁴ L. Pauling, *Nature of the Chemical Bond* (Cornell University Press, Ithaca, 1940).

¹⁵ P. J. Flory, *J. Am. Chem. Soc.* **62**, 1057 (1940).

¹⁶ R. H. Ewell and H. J. Eyring, *J. Chem. Phys.* **5**, 726 (1937). H. Eyring, *ibid.* **9**, 268 (1941). W. Kauzmann and H. Eyring, *J. Am. Chem. Soc.* **62**, 3113 (1940).

to $\pm 0.5^\circ\text{C}$. The results are listed in Table III and plotted in Fig. 9. A straight line is obtained when the logarithm of the viscosity is plotted against the reciprocal of the absolute temperature.

It is apparent from Fig. 9 that the slope of the various lines is not constant but increases with increasing molecular weight or viscosity. The energy barrier calculated from the individual slopes is plotted as a function of the extent of reaction as calculated from the change in iodine number and molecular weight, in Fig. 10, and the data are assembled in Table IV. A linear relationship exists between the extent of reaction and the energy of activation in contrast to the result of Flory, who found the energy of activation independent of the extent of reaction in his work on polyesters. The answer to this discrepancy lies in the fact that Flory prepared only linear polyesters, and consequently, the shape of the molecule would change only in length as the polymerization proceeded. In the case of linseed oil, a highly branched molecule would be produced by polymerization. Since the distance between branches in the linseed oil polymer is of the same order of magnitude as the segment of flow, it is logical that this branched structure would increase the unit of flow.

INTERMEDIATE COMPOUND FORMATION

The results as shown in Figs. 1-6 are in good agreement with the reported changes in properties¹³ of drying oils on heat bodying. Linseed oil and soybean oil give analogous results in contrast to tung oil. Tung oil contains a high proportion of conjugated fatty acids which in a large measure accounts for its greater reactivity. For both soybean oil and linseed oil, the iodine number, hydrogen absorption, and percent soluble in

TABLE III. Temperature coefficient of viscosity of heat bodied linseed oil sample No. II.

Number	Logarithm of viscosity		
	30°C	50°C	75°C
II-4	0.623	0.301	0.041
-5	1.076	0.634	0.256
-6	1.170	0.732	0.322
-7	1.510	1.013	0.544
-8	1.808	1.292	0.792
-9	2.030	1.525	0.977
-10	2.706	2.154	1.846

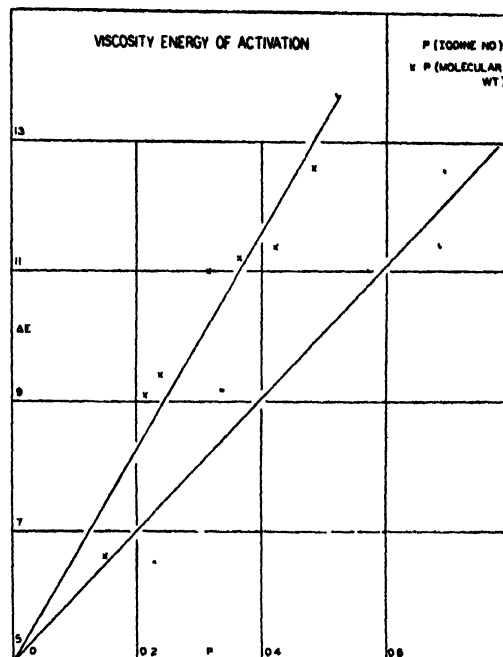


FIG. 10.

acetone decrease rapidly at the beginning of the reaction and tend to approach a constant value at the later stages, while the density and refractive index increase rapidly at the start and then level off. On the other hand, the viscosity and molecular weight experienced relatively little change at the beginning and increased very rapidly at the later stage.

If (p) the extent of reaction is calculated from both the iodine number and molecular weight, different values are obtained from the two properties. To explain this discrepancy the presence of an intrapolymer has been postulated. This idea has been suggested by several other investigators.¹⁷ An intrapolymer consists of either a

TABLE IV. Viscous energy of activation.

Number	Iodine No.	Molecular wt.	ΔE kcal.
II-4	0.223	0.150	6.6
-5	0.346	0.216	9.1
-6	0.454	0.240	9.4
-7	0.570	0.316	11.0
-8	0.631	0.366	11.2
-9	0.688	0.423	11.4
-10	0.697	0.485	12.6

¹⁷ T. F. Bradley, Ind. Eng. Chem. 30, 689 (1938). A. Schwarzman, Oil and Soap 21, 204 (1944).

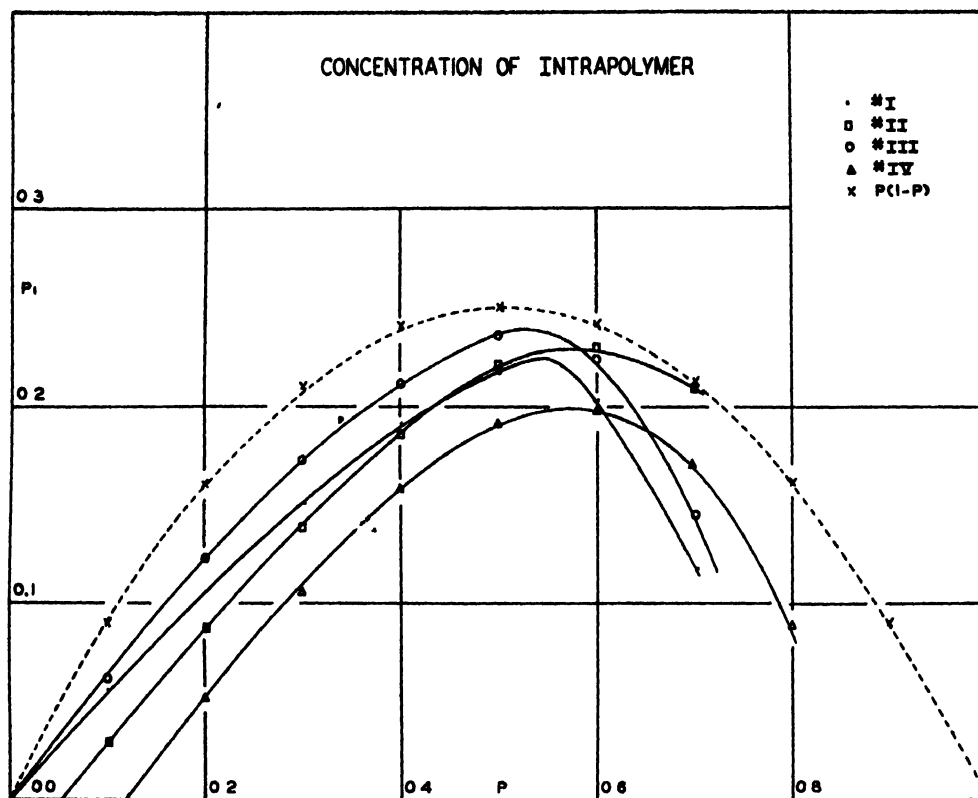


FIG. 11.

monomer glyceride which has two of its fatty acid groups joined together or a polymer that

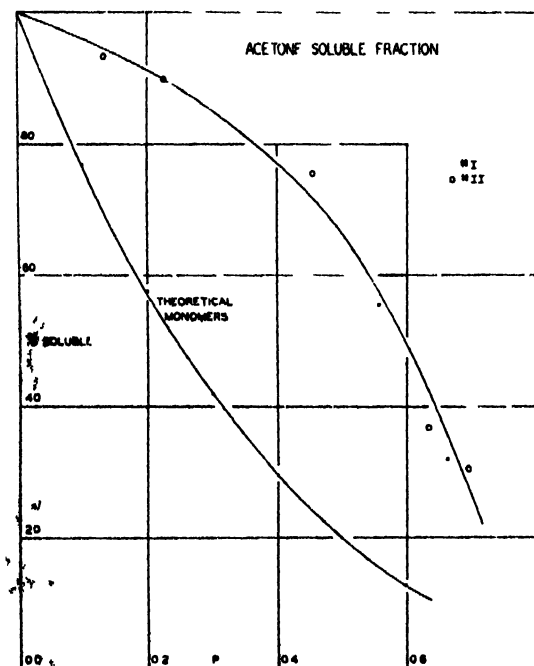


FIG. 12.

contains a ring. Figure 11 shows the magnitude of this difference between the iodine number and molecular weight changes as ordinate and the extent of the reaction as calculated from the iodine number as abscissa. The fact that the concentration of the intrapolymer reaches a maximum at about $p = \frac{1}{2}$, and then decreases, indicates that most of the ring formation is on the monomers. Thus, the intrapolymer is being formed simultaneously with the interpolymer (bonds between fatty acid groups on different glyceride molecules) and at the same time the intrapolymer is continually being changed to an interpolymer by ester interchange which occurs readily at the temperature of heat bodying.

The dotted curve in Fig. 11 is given by the equation $p(1-p)$.

Earlier, the assumption was made that the reactive groups all had the same reactivity, that is, both linolenic and linoleic acids were capable of forming only one bond between the glycerides. However, the presence of trimer and higher fatty acids has been reported in the hydrolyzed products of heat bodied oils. This would indicate that one

of the fatty acids, preferably the linolenic acid, was capable of forming more than one bond which would impart a higher functionality to the oil than used in this work. This effect would reduce the magnitude of the difference between the iodine number and molecular weight but would not change materially the shape of the difference curve.

Figure 12 contains a plot of the percentage of oil soluble in acetone against the extent of the reaction as calculated from the iodine number. This is to be compared with the theoretical amount of monomer predicted by Eq. (13). It is assumed that only monomeric material dissolves in the acetone, which fact has been roughly checked experimentally. The results are vitiated somewhat by the mutual solubility effect which has a tendency to dissolve higher molecular weight material and by the difficulty experienced in effecting a separation between two phases. However, at high extents of reaction the amount of acetone soluble material reaches the weight percent of monomers predicted by the theory. This behavior is in agreement with the postulation of the formation of an intermediate compound having a lower iodine number but the same molecular weight as the monomer.

Stockmayer¹⁸ has measured the gel points of the pentaerythritol-adipic acid system in the presence of various quantities of an inert solvent to prove ring formation. Ring formation should be more prevalent at higher dilutions because of the reduced probability of forming a bond between the different molecules. This should increase the ratio of intrapolymer formation to interpolymer formation. Stockmayer extrapolated the results to a fictitious infinite concentration to obtain a value for the extent of reaction free from ring formation and obtained excellent agreement with the theoretical value.

A sample of linseed oil in 50 percent xylene

¹⁸ W. H. Stockmayer, *Advancing Fronts in Chemistry*, Edited by S. B. Twiss (Reinhold Publishing Corporation, New York, 1945), Vol. 1, p. 70.

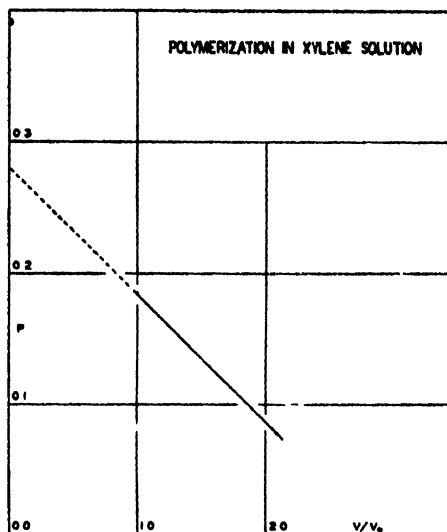


FIG. 13.

solution and a sample of pure linseed oil have been bodied at 300°C. Physical properties of the resultant bodied oils have been obtained. The iodine numbers of the two samples were identical while the molecular weight and viscosity of the oil bodied in solution were less than those for the oil not bodied in solution, which would indicate additional ring formation in the solution.

Figure 13 shows the extent of reaction as calculated from the molecular weight plotted against the ratio (V/V_0) where (V) is the specific volume of the solution and (V_0) the specific volume of the oil. If one assumes that the points fall on a straight line, a value for the extent of reaction at infinite concentration of $p=0.286$ is obtained. This is to be compared with a value of $p=0.392$ obtained from the iodine number. These results may be within experimental error. More data are necessary to determine whether the extrapolation should be a straight line. Additional work is planned to further study this method.

ACKNOWLEDGMENT

The assistance of Miss Helen I. Yeager, who prepared the graphs, is gratefully acknowledged.

Electrostatic and Tensile Properties of Rubber and GR-S at Elevated Temperatures

R. S. HAVENHILL, H. C. O'BRIEN, AND J. J. RANKIN

Research Laboratories, St. Joseph Lead Company, Josephstown, Pennsylvania

This paper describes a new apparatus for measuring electrostatic contact potentials on various materials at elevated temperatures. In this apparatus, the electrostatic charge, acquired by rolling a steel ball down the surface of a rubber test specimen on a heated inclined plane, is measured when the ball drops into the cup of a suitable measuring device such as our electrostatic modulator. With this instrument the contact potentials of both rubber and GR-S was found to become highly negative at elevated temperatures. This apparent "boiling off" of electrons and resultant disruption of electrostatic attractive forces within the material is much greater for GR-S than for rubber and probably accounts for the much greater decrease in tensile of GR-S over rubber at elevated temperatures, and is further confirmation of the electrostatic contact potential theory of reinforcement. By the further application of this theory suitably dispersed compounding materials, which are in effect highly positive at elevated temperatures such as certain proteins, finely divided silica and sodium silicate, have been found to increase substantially the hot tensile strengths of GR-S compounds.

IN a previous paper¹ a new apparatus was described for measuring the electrostatic contact potential of rubber and GR-S compounds, and a correlation of tensile strength and electrostatic contact potential was observed. This work resulted in the formulation of an "Electrostatic Contact Potential Theory of Reinforcement" in which reinforcement is explained on the basis of electrical contact potentials and resultant electrostatic attractive forces set up between the rubber and the reinforcing agents. For example, when a material which has a (+) positive electrostatic charge such as zinc oxide is added to rubber which has a (-) negative charge, strong electrostatic attractive forces are set up between the zinc oxide and the rubber and give rise to reinforcement. Now since (+) and (-) charges tend to neutralize each other, the electrostatic contact potential or electrical charge on the resulting stock should be more positive (less negative) than it was prior to the addition of the zinc oxide. This was found to be the case; and in general, the more positive the stock, the higher the tensile strength.

On the basis of this theory, we would predict that rubber compounds when heated would become highly negative which would indicate a low hot tensile strength. This was found to be the case.

It is the purpose of this paper to describe a new apparatus for measuring contact potentials

at elevated temperatures and to point out that the contact potential of both rubber and GR-S compounds becomes highly negative at 212°F. This increase in negative potential, which may be a "boiling off" of electrons and resultant disruption of electrostatic attractive forces within the material, is much greater for GR-S than for rubber, and probably accounts in part for the much greater decrease in tensile of GR-S over rubber at elevated temperatures and is further confirmation of the electrostatic contact potential theory of reinforcement.

By the application of this theory, materials which maintain their highly positive charge at elevated temperatures should make the GR-S stocks more positive, prevent the release of electrons and increase the hot tensile strengths. This was found to be the case; and certain materials such as proteins, finely divided silica and sodium silicate when added as water dispersions to GR-S latex gave products having considerably improved tensiles at elevated temperatures.

HIGH TEMPERATURE CONTACT POTENTIAL APPARATUS

To measure contact potentials at elevated temperatures, the rubber test specimen must be heated to the desired temperature on an electrically insulated surface, momentarily contacted with a cold reference surface (preferably polished steel), and instantly separated therefrom. The electrical charge or contact potential then can be measured on either the rubber test specimen or

¹ R. S. Havenhill, H. C. O'Brien, and J. J. Rankin, *J. App. Phys.* **15**, 731-740 (1944).

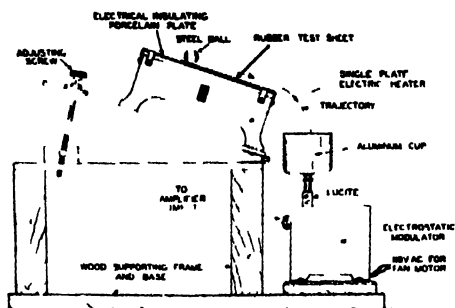
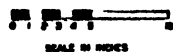


FIG. 1. Rolling ball contact potential apparatus.

on the standard reference contacting surface as the charges are equal and opposite. In our previous apparatus,¹ an insulated steel plunger contacted the test specimen, and the charge was measured on the test specimen which was mounted in a sample holder built into the electrostatic modulator. In this apparatus, it is difficult to build the electrical heater around the sample and provide adequate electrical shielding to prevent the hot a.c. electrical fields around the heater from being picked up on the grid of the audio amplifier. Furthermore, the charged plunger should be out of the test specimen's high temperature electrostatic field when measurements are made. The modulator unit should be isolated from the hot electrostatic field around the sample. To take care of these difficulties, a new apparatus was designed.

ROLLING BALL CONTACT POTENTIAL APPARATUS

In this apparatus (see Fig. 1 and Fig. 2), the electrostatic charge, acquired by rolling a steel ball down the surface of a rubber test specimen on a heated inclined plane, is measured when the ball drops into the cup of a suitable measuring device, such as our electrostatic modulator.

The electrostatic modulator (see Fig. 3) was described completely in our previous paper.¹ Briefly, it consists of a motor-driven fan which cuts the electrostatic field produced by the charged sample and converts it into an alternating-current voltage which easily is amplified and measured by the associated audiofrequency amplifier and built-in output meter. The electrostatic modulator and amplifier were calibrated with an adjustable d.c. high voltage power

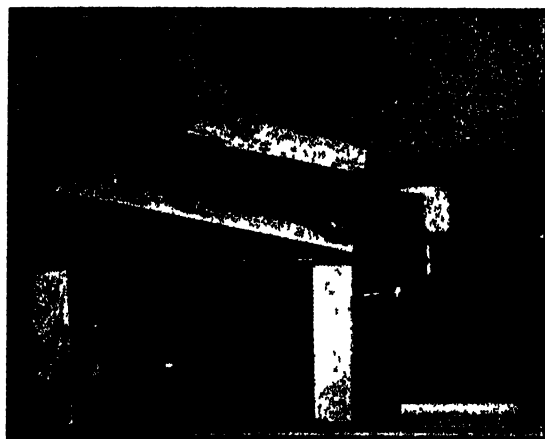


FIG. 2. Rolling ball contact potential apparatus.

supply. The power supply leads were attached to the modulator cup and chassis. Amplifier output meter readings were plotted against input volts from the d.c. power supply. The contact potential values reported are all in terms of these input volts. These contact potential values are not the actual contact potentials since they depend on the capacity of sample, capacity of modulator system, and a number of other factors; however, they are proportional to the actual contact potentials of the rubber test specimens.

With this apparatus, the test specimen is both shielded and electrically insulated from the heater by means of the porcelain enameled steel top plate. The charged steel ball which contacts the specimen is isolated from electromagnetic and electrostatic fields when measurements of its charge are made in the shielded electrostatic modulator unit. While this apparatus has a number of advantages over the plunger type apparatus, for high temperature measurements, it also has some disadvantages. The contact area and velocity of the ball as it leaves the sample are not constant for all samples. For example, a soft stock will have a greater contact area than a hard one and a stock with a tacky surface will retard the velocity of the ball which affects the speed of separation of the ball from the test specimen. Since both of these factors affect the contact potential values, they must be taken into account in testing unknown compounds. For similar compounds, these differences usually do not exceed 10 percent error. The relative contact area can be estimated approximately by

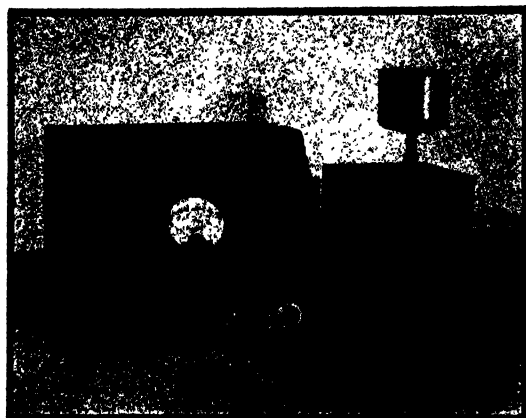


FIG. 3. Electrostatic modulator with audio amplifier and output meter.

coating the steel ball with "fingerprint" ink and measuring the length and width of the black streak thus produced. The relative velocity of the ball as it leaves the sample can be estimated from its trajectory. The greater the velocity of the ball, the greater will be the horizontal range on the trajectory curve. This distance can be easily measured from the imprint of the ball on carbon over white paper.

MECHANICS OF ROLLING BALL CONTACT POTENTIAL APPARATUS

The mechanics of the rolling ball apparatus along with the derivation of the formula relating horizontal range (S) to velocity (V) are of theoretical interest and have been calculated by L. E. Carlson of our Research Laboratories as follows. Refer to Fig. 4.

1. Total Area of Contact

Length of Streak = 6 inches
Width of Streak = $1/16$ inch
Area = $6 \times 1/16 = .375$ sq. in.

2. Pressure

Weight of ball = .0544 lb.
Diameter of ball = $23/32'' = .718$ inch
Diameter of contact = $1/16''$
Instantaneous area of contact = $(1/32)^2 \pi$
Pressure = $.0544 / (1/32)^2 \pi = 17.8$ lb./sq. in.

3. Velocity

Assume that potential energy of ball at top of incline is converted into rotational and trans-

lational kinetic energy by the time it reaches the bottom of the incline of height (h). Thus:

$$mgh = \frac{1}{2}mV^2 + \frac{1}{2}IW^2.$$

Substituting values of I and W the energy equation reduces to

$$mgh = \frac{1}{2}mV^2 + \frac{1}{2}mV^2,$$

$$V = \left(\frac{10gh}{7} \right)^{\frac{1}{2}},$$

$$V = \left(10 \times 32.2 \times \frac{.104}{7} \right)^{\frac{1}{2}} = 2.19 \text{ ft./sec.}$$

m = mass of ball,

h = height of incline (.104'),

g = acceleration of gravity,

V = translational velocity (at bottom of incline) parallel to it,

I = moment of inertia,

$= \frac{1}{2}mr^2$ where r is radius of ball,

W = rotational velocity

$= V/r.$

V is the maximum velocity obtainable, assuming the ball rolls and does not slide. Actually, tackiness of the surface and indentation of the surface slow the ball somewhat.

The following is a means of determining actual velocity from horizontal range which is the horizontal distance the ball moves while in the air.

RELATION OF VELOCITY TO DISTANCE (S)

Assuming the ball starts the trajectory path with a velocity V and travels the horizontal distance (S) and the vertical distance .359 ft. in time t , the following equations of motion may be set up.

$$.359 = \frac{1}{2}gt^2 + V \sin 12^\circ t,$$

$$S = V \cos 12^\circ t.$$

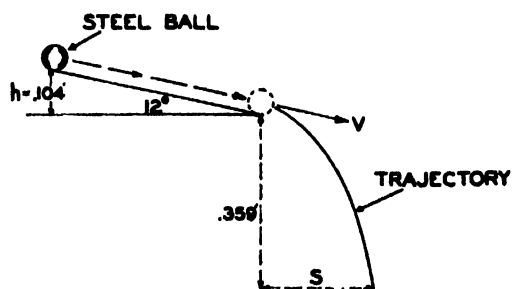


FIG. 4. Mechanics diagram of rolling ball apparatus.

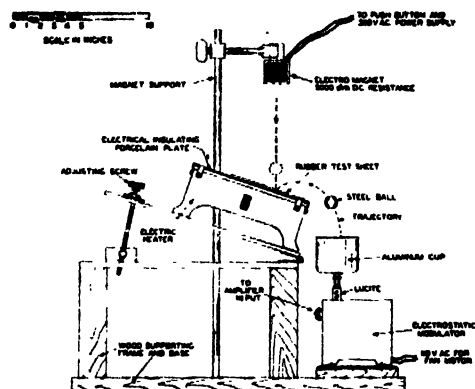


FIG. 5. Bouncing ball contact potential apparatus.

Eliminating t and solving for V ,

$$V = \frac{4.11 \times S}{(.359 - .21265S)^{\frac{1}{2}}} \text{ ft./sec.}$$

Since S can be measured experimentally, V can be calculated using the above formula.

BOUNCING BALL CONTACT POTENTIAL APPARATUS

In order to secure a more intimate contact between the steel ball and the test specimen, the "bouncing ball" type apparatus was designed. Refer to Fig. 5. In this apparatus, which is in part similar to the rebound apparatus described in a previous paper,² the steel ball is held above the test specimen by means of an electromagnet and is dropped upon the rubber test specimen which is mounted on a heated inclined plane and the ball bounced into the electrostatic modulator where its charge is measured. With this apparatus, the work done on the sample during a test is so great that there is some evidence that it brings about a structural change in the rubber. In any case, repeated tests on the same spot in the rubber show a progressive decrease in charge (stocks become more positive) which would indicate an increase in tensile strength.

MECHANICS OF BOUNCING BALL CONTACT POTENTIAL APPARATUS

The mechanics of the bouncing ball apparatus are of theoretical interest and are given as follows. Refer to Fig. 6.

² R. S. Havenhill and J. J. Rankin, Ind. Rubber World 107, 365-368 (1943).

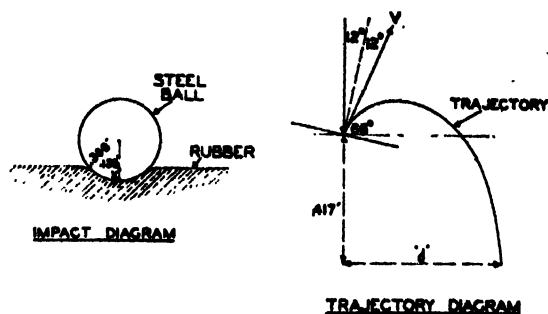


FIG. 6. Mechanics diagram of bouncing ball apparatus.

1. Area of Contact

Diameter of contact spot = .25 inch.

Determined experimentally by coating ball with fingerprint ink.

Area = $(.125)^2 \pi = .049$ sq. in.

2. Pressure

Diameter of ball = .718 inch.

Weight of ball = .0544 lb.

Diameter of contact spot = .25 inch.

Depth of penetration = X .

$X = .359 - [(.359)^2 - (.125)^2]^{\frac{1}{2}}$

$X = .022$ inch.

The ball falls 7.25 inches on to the rubber. It penetrates the surface a depth X . From the law of conservation of energy, it can be shown that the weight of the ball times the distance it falls equals the force exerted on the rubber times the depth of penetration X . Thus:

$$.0544 \times 7.25 = F \times .022.$$

$$F = 17.9 \text{ lb.}$$

$$\text{Pressure} = \frac{17.9}{.049} = 365 \text{ lb./sq. in.}$$

3. Velocity

If the rubber stocks were 100 percent resilient, the velocity of the ball leaving the stock would be equal in magnitude to the velocity of the ball falling 7.25 inches from rest.

$$V = (2gS)^{\frac{1}{2}} = \left(\frac{2 \times 32.2 \times 7.25}{12} \right)^{\frac{1}{2}},$$

$$V = 6.24 \text{ ft./sec.}$$

This is the maximum velocity obtainable. Since resilience is proportional to kinetic energy regained after rebound, and kinetic energy is proportional to velocity squared, it can be shown that actual velocity after rebound is as follows:

$$V = .624R^{\frac{1}{2}} \text{ ft./sec.},$$

where R is resilience in percent.

Another means of measuring this velocity is by the use of d , the horizontal distance the ball moves after bouncing. If the ball starts with a velocity V (as shown in Fig. 6) and falls a distance of 0.417 ft. in time t , equations of motion can be set up.

Assuming the angle of impact equals the angle of rebound, these equations are

$$\begin{aligned} -V \sin 66^\circ t + \frac{1}{2}gt^2 &= 0.417, \\ V \cos 66^\circ t &= d. \end{aligned}$$

Solving for V in terms of d

$$V = \frac{10.1d}{(.417 + 2.246d)^{\frac{1}{2}}} \text{ ft./sec.}$$

Values of V obtained from this formula agreed with values obtained from the resiliency-velocity formula within about 5 percent. Hence, the assumption of equal angles has experimental justification.

TESTING PROCEDURE

A 6"×6"×.075" tensile sheet³ cured in a clean chromium-plated mold is used for test. The carefully prepared specimen,¹ whose surface should be free from dust, lint, surface bloom, or other foreign matter and free from electrical charge, is placed on the inclined plane and brought to the desired testing temperature by proper adjustment of the Variac supplying the current for the electric heater. Temperature can be measured by means of a thermocouple placed on the stock and another built into the top surface of the inclined plane. A four-minute heating period is usually required to bring the tensile sheet to a testing temperature of 212°F. The steel ball (ballbearing) which is used to contact the sample is first cleaned with c.p. acetone. The ball is then picked up and gently

placed on top of the rubber test sheet and allowed to roll freely down the rubber on the incline and drop into the cup of the electrostatic modulator where its charge is measured on the output meter of the audio amplifier. This test is repeated several times, taking care to roll the ball over a fresh uncontacted surface each time. The test specimen should be placed slightly over the lower edge of the inclined plane so that arcing and resultant discharge of the ball cannot take place between the ball and the metal of the heater.

The specimen preparation and test procedure for the "bouncing ball" apparatus are the same as for the "rolling ball" apparatus with the exception that in the former, the ball is released from an electromagnet and bounced instead of rolled into the electrostatic modulator. A thicker test specimen is sometimes desirable with this apparatus, and samples 0.20 inch in thickness were used for certain tests. All contact potential measurements must be made in a constant temperature, constant humidity room; 78°F and 45 percent relative humidity testing conditions have been found to be satisfactory. Like all electrostatic work, results are difficult to duplicate, and a large number of tests must be run.

EXPERIMENTAL DATA

Increase of Contact Potential with Temperature

Our electrostatic contact potential theory of reinforcement predicts that GR-S should become highly negative at elevated temperatures in order to explain the low hot tensiles of GR-S. To check this, contact potentials were run at 78°F and 212°F on a 20-volume zinc oxide loaded GR-S stock and also on a similar rubber stock. Data were obtained with both the "rolling ball" and "bouncing ball" contact potential equipment. Rebound data were also obtained using the St. Joe Inclined Plane Falling Ball Rebound Tester.² Hot tensile data were obtained using the St. Joe Electric Heater Attachment⁴ for the Scott Tensile Machine (see Fig. 7). These data are shown in Table I. The zinc oxide used in these and all subsequent tests was a

⁴ Test Method Card No. 66 File S-2 dated 10/23/43—Research Compounding Section, Research and Development Division of the Office of Rubber Director, Washington, D. C.

³ A.S.T.M. D15-41.

TABLE I. Comparison of rolling ball and bouncing ball electrostatic contact potentials of rubber and GR-S.

Test	GR-S Formula A		Rubber Formula B	
	78°F	212°F	78°F	212°F
*Electrostatic contact potential volts				
1. Rolling ball apparatus	116	293	103	140
2. Bouncing ball apparatus	36	89	28	37
Tensile lb./sq. in.	1000	115	3500	2100
% elongation	930	400	620	625
Modulus @ 400 percent (lb./sq.in.)	125	115	1200	—
% rebound	35	36	44	56
Shore hardness	42	—	51	—
Rex hardness	44	—	53	—

Formulas

A		B	
GR-S	100	Rubber	100
Sulfur	2.5	Sulfur	3.5
Captax	1.5	Captax	0.6
ZnO	113.	Stearic acid	1.5
	—	ZnO	113.
	217.0		218.6

Cure 55'×280°F

Cure 35'×280°F

* The contact potential values are in terms of negative input volts at modulator. The smaller the value, the more positive (+) the potential.

fine particle size (0.2 micron average diameter), fast-curing type produced by the St. Joe Electrothermic process. It will be noted that the GR-S contact potential increased about 150 percent while the rubber contact potential increased less than 50 percent. This was true for both the "rolling" and "bouncing" ball type apparatus. These data checked the theory, and it was decided to repeat these tests and run additional



FIG. 7. St. Joe hot tensile apparatus.

ones at different temperatures on both high zinc and pure gum rubber and GR-S stocks. Since the "bouncing ball" apparatus gave such low contact potentials because of small contact area, the results were difficult to check and all subsequent tests were run using only the "rolling ball" apparatus. These contact potential data are shown in Table II. The tensile data are shown in Table III. It will be noted that in all cases there is an increase in negative contact potential as the temperature is increased. This corresponds with a decrease in tensile with temperature and is further confirmation of the electrostatic contact potential theory of reinforcement.

It is interesting to note that the high contact potential values observed at elevated tempera-

TABLE II. Effect of temperature on electrostatic contact potentials of rubber and GR-S.

Formula	Composition	Min. cure @ 280°F	Electrostatic contact potentials rolling ball apparatus volts				% rebound		Hardness 78°F	
			78°F	145°F	212°F	278°F	78°F	212°F	Shore max.	Rex
C	Pure gum rubber	35	100	120	150	250	48	59	42	43
D	Pure gum GR-S	55	150	225	325	450	42	43	35	35
B	20-vol. ZnO rubber	35	85	115	140	185	44	56	51	53
A	20-vol. ZnO GR-S	55	115	200	285	400	35	36	42	44

Formulas

C		D	
Rubber	100	GR-S	100
Sulfur	3.5	Sulfur	2.5
Captax	0.6	Captax	1.5
Stearic acid	1.5	Zinc oxide	5
Zinc oxide	5		109.0
	110.6		

See Table I for Formulas A and B.

TABLE III. Effect of temperature on tensile of rubber and GR-S compounds.

Formula	Composition	Cure min. @ 280°F	78°F			145°F			212°F			270°F		
			Tens. lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.	Tens. lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.	Tens. lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.	Tens. lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.
C	Pure gum rubber	35	3350	775	315	2850	675	400	2200	760	275	160	250	—
D	Pure gum GR-S	55	145	450	75	85	250	—	50	150	—	25	160	—
B	20-vol. ZnO rubber	35	3500	620	1200	2850	640	750	2200	625	800	485	400	465
A	20-vol. ZnO GR-S	55	1000	930	125	225	500	150	125	425	100	65	275	—

See Tables I and II for formulas.

tures tend to persist for a short time after the stock is cooled, indicating an alignment or orientation of charges such as a semi-permanent polarization of the stock.

ADDITION OF POSITIVE (+) MATERIALS TO GR-S TO INCREASE HOT TENSILE STRENGTHS

A number of materials having, in effect, a less negative (more +) electrostatic contact potential were added to GR-S to make it less negative (more +) at elevated temperatures and, according to the theory, increase its tensile strength. Some of these materials, such as Flectol H, while (+) and effective in increasing tensile strength at room temperatures, became highly negative at elevated temperatures; and for this reason, neither made the stocks more positive nor increased the tensile at elevated temperatures. A few of these materials when properly dispersed maintained their positive charge at elevated temperatures, making the GR-S less negative and gave increased tensile strength stocks. These materials were finely divided SiO_2 , Na_2SiO_3 , and certain proteins such as egg white. These materials were not effective when milled into GR-S and did not change appreciably the contact potentials or tensile strengths. A careful examination of the stocks showed poor dis-

persion of all of these materials. In order to get effective dispersions, these materials were dispersed in water and the water dispersions added to GR-S latex. The procedure was as follows:

LATEX DISPERSIONS

To 1000 grams of Type II GR-S latex (30 percent solids) were added water dispersions of the positive (+) materials in such an amount as to give 20 percent on the GR-S solids. Then, 600 grams of 50 percent Black Label No. 20 zinc oxide dispersed in water with 1 percent of Stablex B were added. Coagulation was carried out by the addition of 100 to 200 cc of 28 percent Epsom Salt solution. Egg white and, especially, sodium silicate are good dispersing agents for zinc oxide and act as stabilizing agents, making it difficult to coagulate the latex. After coagulation, the material is dried at 190°F for approximately twenty-four hours or until the moisture content is under .35 percent. The resulting GR-S master batch contains approximately 50 percent of zinc oxide and 20 percent (on GR-S) of the added (+) material. Compounding then was carried out and regular milling procedure followed in basic Formula E. Formula E is similar to Formula A, except that

TABLE IV. Addition of (+) materials to GR-S latex.

Composition	Cure min. @ 280°F	Electrostatic contact potentials volts		Original tensiles 78°F			Tensiles @ 212°F			% rebound		Hardness 78°F shore 15 sec.
		78°F	212°F	Tens. lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.	Tens. lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.	78°F	212°F	
Standard	50	100 v	275 v	900	850	225	150	300	—	36	37	40
20% egg white	20	93 v	180 v	1475	538	1000	400	250	—	37	40	49
20% SiO_2	50	95 v	195 v	1325	990	200	300	400	300	33	34	37
20% Na_2SiO_3	40	90 v	175 v	1475	450	1200	400	200	—	36	39	49

Positive (+) materials added
2 egg whites added to 1000 grams of Type II GR-S latex (30% solids)
400 grams of 15% santocel (SiO_2) dispersed in water
120 grams of 50% sodium silicate in water.
Basic Formula E (See Text).

Formula E.

Parts by weight	
50 percent zinc oxide master batch from GR-S latex (spec.)	200
Sulfur	2.5
Captax	1.5
	204.0

it contains 13 volumes of zinc oxide instead of 20 volumes.

The tensile, rebound, and contact potential data are shown in Table IV. It will be noted that these materials markedly increased the contact potential (made it more positive) and increased both the room-temperature tensile and the hot tensile of GR-S. The hot tensile is more than doubled in case of the egg white (albumen) and sodium silicate stocks. Modulus, hardness, and rate of cure also are increased while rebound data are of the same order. The stocks containing the egg albumen were different from any of the other stocks in that they milled out extremely smoothly and gave a long, plastic, uncured sheet when taken from the mill. They resembled natural rubber more in this respect than any other GR-S material we have observed.

DISCUSSION OF RESULTS

In order to show the relations between temperature, contact potential, and tensile, additional data were obtained at various temperatures on the standard latex compound containing no sodium silicate and on one containing 20 percent sodium silicate. These data are shown in Table V and have been plotted in the form of curves. Refer to Figs. 8 and 9. Figure 8 shows the increase in contact potential with tempera-

ture and also that the addition of Na_2SiO_3 greatly decreases the contact potential and makes the stock more positive (+). Figure 9 shows how the tensile drops off with temperature and shows the superiority of the Na_2SiO_3 stock.

In Fig. 10, the contact potential data for all the rubber compounds run at different temperatures have been plotted against the tensile strength values; and it appears that there is a relation between tensile strength and contact potential.

In Fig. 11, the contact potential data for all the GR-S and GR-S latex compounds run at different temperatures have been plotted against the corresponding tensile strength values and again it appears that there is a correlation of contact potential and tensile strength; however, it is a different one from that for rubber as might be expected from the difference in physical and chemical properties of the two materials. These data are all further confirmation of the electrostatic contact potential theory of reinforcement.

SUMMARY AND CONCLUSIONS

A new apparatus has been described for measuring contact potentials of rubber and GR-S compounds at elevated temperatures. In this apparatus, the electrostatic charge, acquired by rolling a steel ball down the surface of a rubber test specimen on a heated inclined plane, is measured when the ball drops into the cup of the electrostatic modulator. This potential, while not the actual contact potential, is, nevertheless, proportional to it. With this apparatus, the contact potential of GR-S at elevated temperatures was found to increase much more (become more negative) than for rubber. The release of electrons (increase in negative contact potential)

TABLE V. Effect of temperature on GR-S containing sodium silicate.

Temperature of test	Contact potential volts		Standard stock no sodium silicate			20% sodium silicate		
	Standard stock	20% Na_2SiO_3 stock	Tensile lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.	Tensile lb./sq. in.	Elong. %	Mod. @ 400% lb./sq. in.
78°F	100	85	900	850	225	1475	475	1000
145°F	180	125	500	400	300	890	425	700
212°F	275	185	150	300	—	400	200	—
279°F	410	255	50	150	—	275	225	—

Basic Formula E.

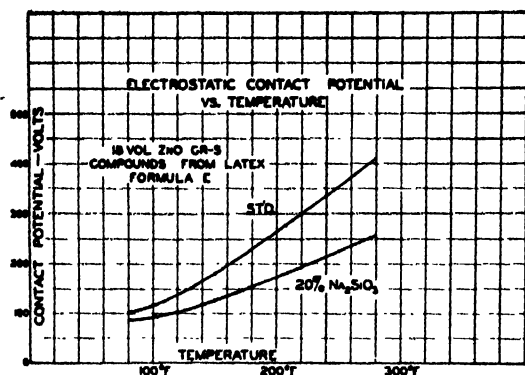


FIG. 8. Contact potential vs. temperature for sodium silicate stocks.

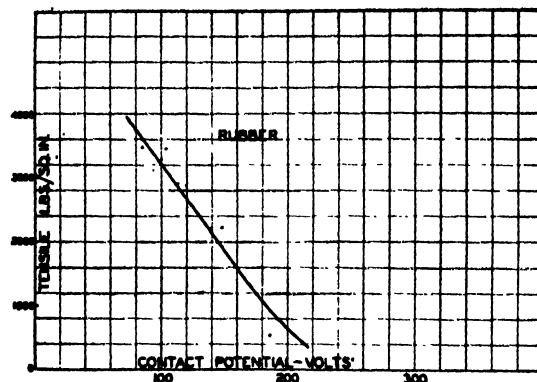


FIG. 10. Contact potential vs. tensile strength for all rubber compounds.

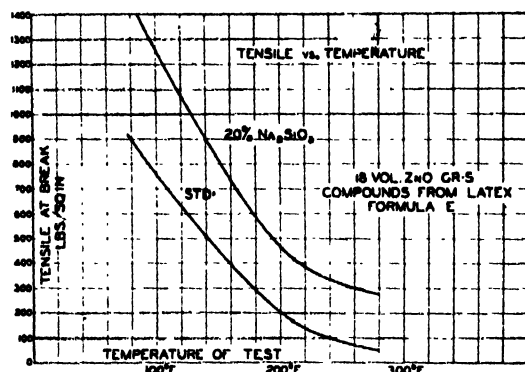


FIG. 9. Tensile vs. temperature for sodium silicate stocks.

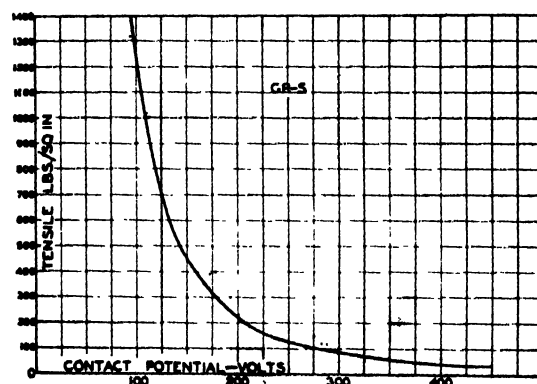


FIG. 11. Contact potential vs. tensile strength for all GR-S compounds.

and consequent disruption of electrostatic attractive forces within the material at elevated temperature probably partly accounts for the much greater decrease in tensile of GR-S over rubber and is further confirmation of the electrostatic contact potential theory of reinforcement.

By the further application of this theory, highly positive (+) materials, such as certain proteins, finely divided silica, and sodium silicate, which retain their positive charge at elevated temperatures and make the stocks more positive, have been found to more than double the hot

tensiles of high zinc oxide loaded compounds made from GR-S latex.

ACKNOWLEDGMENT

The authors wish to acknowledge the valuable assistance of the Staff of the St. Joe Research Laboratory and particularly that of Mr. G. F. Weaton, Plant Manager, Mrs. D. E. Taylor, Harold Emery, L. E. Carlson, L. A. Cha, and Charles Black, who have contributed to the preparation of this paper.

Significance of the Equation of State for Rubber*

EUGENE GUTH, PAUL E. WACK, AND ROBERT L. ANTHONY
Polymer Physics Laboratory, University of Notre Dame, Notre Dame, Indiana

The physical significance of stress-strain curves and of isometrics obtained by the relaxation method is discussed and clarified. Stress-strain curves taken at various temperatures give the correct dependence of stress upon temperature if they are taken sufficiently fast so that stress relaxation does not mask the temperature dependence. Isometrics obtained after previous relaxation of the sample are shown to depend upon duration and temperature of the relaxation by a numerical factor only. The basis for this behavior is the factorization of the stress into a factor depending upon extension and temperature only which corresponds to the equation of state and another factor depending upon the temperature T^* and the duration of the relaxation process. For simple stress relaxation, the same factorization holds with T^* equal to T . A general theory is formulated for time dependent elastic phenomena by generalizing Boltzmann's theory. The theory explains why factorization does not hold for creep, in agreement with experiment.

1. INTRODUCTION

THE physical significance and the molecular interpretation of equations of state for gases, liquids, and solids are well known. The equation of state for a gas may be given in the form

$$p = p(V, T), \quad (1)$$

where p is the pressure, V the volume, and T the absolute temperature. The equation of state for a solid, which is deformed unilaterally, may be obtained from Eq. (1) by replacing p by $(-Z)$, the (nominal) stress (=force per unit original cross section) and V by ϵ , the strain (=extended length divided by original length minus one)

$$Z = Z(\epsilon, T). \quad (2)$$

A stress-strain curve

$$Z = Z(\epsilon) \quad (2a)$$

is the *isothermal* obtained from Eq. (2) at constant temperature T . A stress-temperature curve

$$Z = Z(T) \quad (2b)$$

is the *isometric* obtained from Eq. (2) at a constant extension ϵ . An extension-temperature curve

$$\epsilon = \epsilon(T) \quad (2c)$$

is the *isotonic* obtained from Eq. (2) at a constant stress Z . Equations (2a), (2b), and (2c) are, of course, equivalent. For certain solids, the iso-

thermals, or the isometrics or the isotonics may be obtained experimentally in a direct way.

For rubber, none of the "iso"-curves may be obtained directly. For instance time is needed to take an isothermal, a stress-strain curve. During this time the stress relaxes by an amount depending upon the conditions under which the curve is taken. The purpose of the present paper is to clarify the role of the time element in the various methods for obtaining the equation of state.

2. STRESS-STRAIN CURVES

Apparently the simplest way of obtaining the equations of state consists in the taking of stress-strain curves at various temperatures. Unfortunately, the time element needed for this procedure may mask or even reverse the correct dependence of stress upon temperature. Stress-

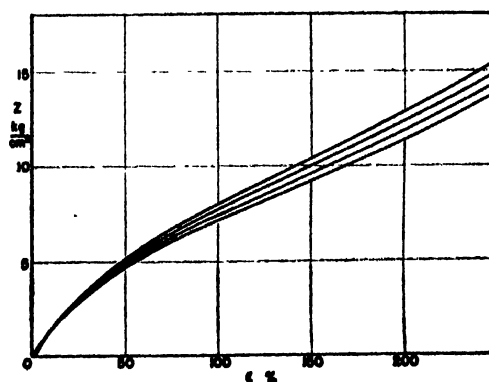


FIG. 1. Stress-strain curves for various temperatures.

* This work was done in 1943, but its publication had to be delayed.

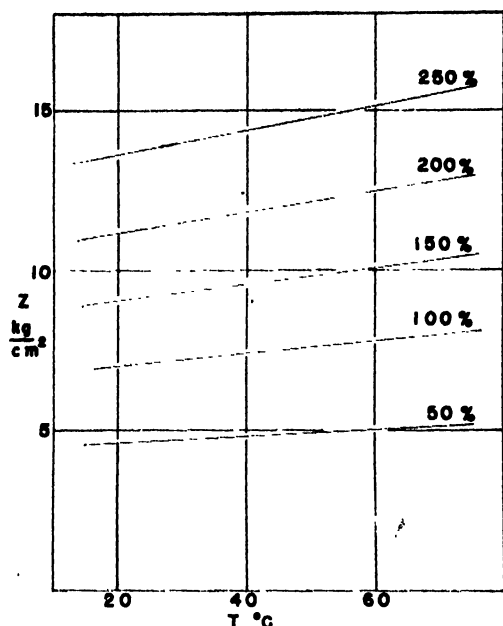


FIG. 2. Isometrics derived from Fig. 1.

strain curves taken at *slow* constant rates of loading or stretching yield on cross-plotting at constant strain a decrease of the stress with increasing temperature, in contradiction to the kinetic theory of rubber electricity.¹

For *fast* stress-strain curves the correct dependence (i.e., increase) of the stress upon the temperature is obtained. Figure 1 shows stress-strain curves at various temperatures obtained with the fast stress-strain apparatus,² developed in the Polymer Physics Laboratory at Notre Dame. Figure 2 shows isometrics which resulted from Fig. 1 on cross-plotting; they exhibit the increase of stress with increasing temperature.

This method alone, however, cannot give a reliable equation of state. For, it is not *a priori* known *how fast*³ the stress-strain curve has to be taken in order to give the "correct" dependence of the stress upon the temperature. In addition, the adjective "correct" needs further elaboration. At any rate this method would have to be

¹ H. M. James and E. Guth, J. Chem. Phys. 11, 455 (1943).

² S. L. Dart, R. L. Anthony, and P. E. Wack, Rev. Sci. Inst. 17, 106 (1946).

³ Because of the rise of temperature on fast stretching these stress-strain curves are adiabatics in reality. However, the difference between adiabatics and isothermals is within the experimental error for extensions up to crystallization.

supplemented by a study of the effect of the speed of stretching upon the stress-strain curve.

3. ISOMETRICS BY RELAXATION METHOD

The relaxation method of obtaining isometrics operates as follows: A sample is kept at constant extension at a constant temperature T^* for the time interval t^* . The stress relaxes as shown in Fig. 3. Then the temperature is lowered and the change in stress measured, giving the desired isometrics as indicated in Fig. 4.

The time t^* is chosen so that in the time interval $(t^{**} - t^*)$ (cf. Fig. 3) needed to take the measurements over the temperature interval $T^* - T^{**}$ (cf. Fig. 4) the amount of stress relaxation is within the experimental error.

In earlier work it was customary to take measurements from T^{**} back to T^* also and call the isometrics reversible or equilibrium curves, if

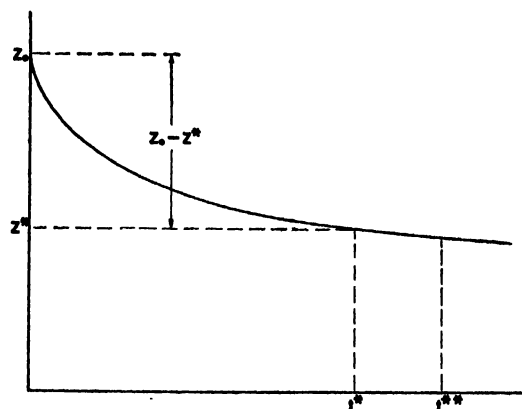


FIG. 3. Illustration of the relaxation method for obtaining isometrics, Z_0 is the initial stress at $t=0$.

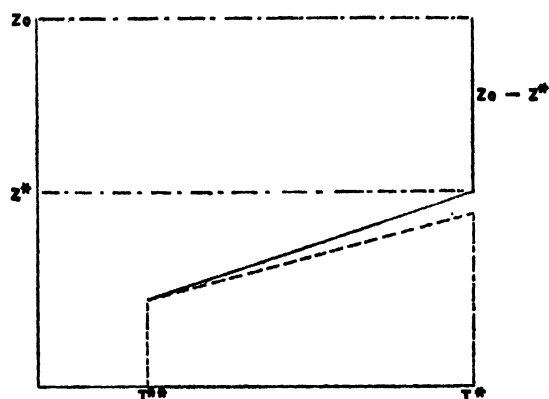


FIG. 4. Isometric obtained after pre-relaxation according to Fig. 3.

the branch coming back coincided with that going down. This is unnecessary, however. Suppose one has chosen t^* so that no measurable stress relaxation takes place going down from T^* to T^{**} , that is, for the time $(t^{**} - t^*)$, but that stress relaxation is not negligible for the interval $2(t^{**} - t^*)$ needed to cover also the return T^{**} to T^* . Then, when coming back, one would obtain the dotted line in Fig. 4. For thermodynamic computations one still may use the branch on going down. This statement will be obvious after considering the theory of the relaxation method of obtaining isometrics in the following sections of this paper.

Apart from the clarification of this item, there were other important points left somewhat obscure. In the actual experiment all rubbers show some (permanent or temporary) set, when the stress is released, after the isometrics are taken. Should one take this *final* unstretched length or the *initial* one (before the experiment started) in computing the extension ϵ ? Meyer and Ferri,⁴ for instance, have used the final length. Our theory will show that the initial length should be taken instead. In previous work (Meyer and Ferri,⁴ our own work,⁵ and Wood and Roth⁶) the role of the temperature T^* and of the duration t^* of the pre-relaxation was left somewhat unclear. Also the question of the *absolute value* of the stress was not discussed.

4. STRESS RELAXATION

The importance of the pre-relaxation in the relaxation method for isometrics necessitates a closer study of the laws of stress relaxation.

Our experimental work on the stress relaxation of natural and synthetic gum and tread stocks may be summarized in the factorization relation

$$Z(\epsilon, T; t) = F(\epsilon, T)G(t, T). \quad (3)$$

Stress relaxation differs from the equation of state in the entering of the time as a new variable. This means, for instance, that it is not possible to

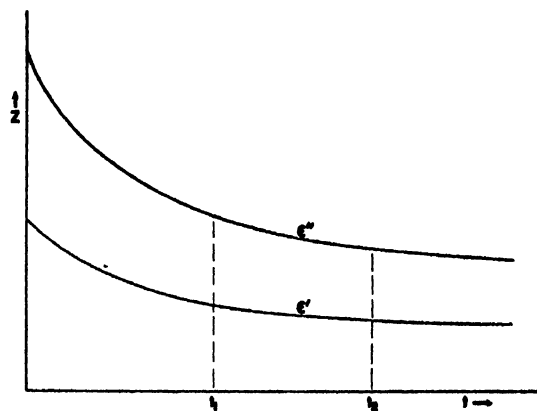


FIG. 5a. Similarity of stress-time curves.

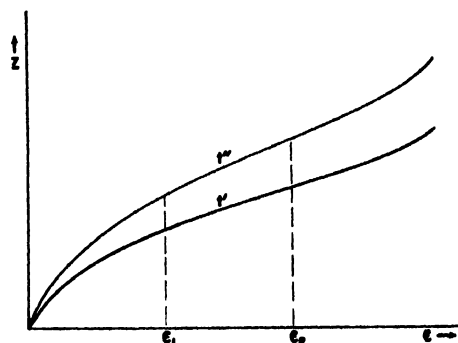


FIG. 5b. Similarity of stress-strain curves obtained from stress-time curves of the type shown in Fig. 5a.

derive for all times t the laws of creep ($Z = \text{const.}$) from Eq. (3) by solving for ϵ .

The Eq. (3) implies the constancy of the following ratios:

$$\begin{aligned} \frac{Z(\epsilon'', T; t)}{Z(\epsilon', T; t)} &= \frac{Z(\epsilon'', T; t_2)}{Z(\epsilon', T; t_2)} \\ &= \dots = \frac{F(\epsilon'', T)}{F(\epsilon', T)}, \quad (4a) \end{aligned}$$

$$\begin{aligned} \frac{Z(\epsilon_1, T; t'')}{Z(\epsilon_1, T; t')} &= \frac{Z(\epsilon_2, T; t'')}{Z(\epsilon_2, T; t')} \\ &= \dots = \frac{G(t'', T)}{G(t', T)}. \quad (4b) \end{aligned}$$

The meaning of the notation in Eqs. (4a) and (4b) is made clear in Figs. 5a and 5b. These figures exhibit the geometric implications of the factorization. The relaxation ($Z-t$) curves for various constant extensions ϵ and the stress-

⁴ K. H. Meyer and C. Ferri, *Helv. Chim. Acta* **18**, 570 (1935). Translated in *Rubber Chem. Tech.* **8**, 319 (1935).

⁵ R. L. Anthony, R. H. Caston, and E. Guth, *J. Phys. Chem.* **46**, 826 (1942). Reprinted in *Rubber Chem. Tech.* **16**, 297 (1943).

⁶ L. A. Wood and F. L. Roth, *J. App. Phys.* **15**, 781 (1944).

strain ($Z-\epsilon$) curves⁷ for various constant times t' , t'' , . . . are *similar*, i.e., they may all be superposed on each other by multiplication by constant factors.

5. GENERAL THEORY OF TIME ELASTICITY

Time elasticity shall designate any process in which stresses cause deformations with the time element as an important factor. A general formalism may be given in agreement with the factorization (Eq. (3)) by a generalization of Boltzmann's theory of time-elastic processes.

The fundamental integral equation is

$$Z(t) = F[\epsilon(t)] - \beta \int_{-\infty}^t \varphi(t-t') F[\epsilon(t')] dt' \quad (5)$$

β is a (small) constant. $\beta\varphi(t-t')$ is the memory function of Boltzmann. Equation (5) reduces to Boltzmann's equation, if the instantaneous stress-strain relation

$$Z = F(\epsilon) \quad (6a)$$

reduces to Hooke's law

$$Z = E\epsilon, \quad (6b)$$

where the constant E denotes Young's modulus.

For *relaxation*:

$$\epsilon(t) = \epsilon(t') = \begin{cases} 0, & t < 0 \\ \epsilon = \text{const.}, & t > 0. \end{cases}$$

Equation (5) reduces to

$$Z(t, \epsilon) = F(\epsilon)[1 - \beta\psi(t)] = F(\epsilon)G(t), \quad (7)$$

where

$$\psi(t) = \int_0^t \varphi(t-t') dt'. \quad (8)$$

Equation (7) is identical with Eq. (3). (For simplicity we did not indicate in Eq. (5) the temperature as another variable.)

For *creep* we have to invert our fundamental Eq. (5). For short times this is simple. Then the term with β will be small and terms of order β^2 or

⁷ These stress-strain curves are obtained by cross-plotting from the relaxation curves (Fig. 5a). They differ from stress-strain curves obtained at a constant rate of loading or stretching.

higher may be neglected. We obtain

$$\epsilon(t) = F^{-1} \left\{ Z(t) + \beta \int_{-\infty}^t \varphi(t-t') Z(t') dt' \right\} \quad (9)$$

and for creep

$$\begin{aligned} Z(t) = Z(t') &= \begin{cases} 0, & t < 0 \\ Z = \text{const.}, & t > 0, \end{cases} \\ \epsilon(t) &= F^{-1} \{ Z[1 + \beta\psi(t)] \}. \end{aligned} \quad (10)$$

Equation (10) follows from Eq. (7) by solving for ϵ and yields factorization for creep in the special case of Hooke's law, where $\epsilon = F^{-1}(Z) = Z/E$ but not in the general case of Eq. (6a). We notice that had we assumed factorization for creep, we would have arrived at an equation different from (9) which would have contradicted Eq. (5). Our experiments show that—in contrast to relaxation—factorization does not hold for creep.

All time-elastic phenomena may be derived from the fundamental Eq. (5). For instance, we can obtain the dependence upon time of stress-strain curves obtained at constant rate of loading or of stretching.⁸ The *fast* processes of forced or free vibrations may be treated using the same equation as a general frame.

The determination of the memory function from molecular models we shall defer to another paper. The function $Z = F(\epsilon)$ is the stress-strain relation as given by the statistical theory of rubber elasticity.¹

$$Z = K \cdot T \left[1 + \epsilon - \frac{1}{(1 + \epsilon)^2} \right]. \quad (11)$$

For *shear* deformations Hooke's law (6b) holds (with ϵ as shear-strain) up to moderate extensions and Eq. (5) reduces to Boltzmann's. For shear there is then factorization for *both* creep and relaxation.

6. THEORY OF ISOMETRICS FROM RELAXATION METHOD

Experimental work at Notre Dame on natural and synthetic gum stock may be summarized in the relation

$$Z(\epsilon, T; T^*, t^*) = F(\epsilon, T), G(T^*, t^*). \quad (12)$$

⁸ Cf. G. Slomimsky, Acta Phys. Chem. U.S.S.R. 12, 99 (1940), for the case of Boltzmann's theory.

Here T^* and t^* designate the temperature and duration of the pre-relaxation process which precedes the taking of the isometrics (cf. Figs. 3 and 4). Equation (12) expresses a generalized factorization. Again, as in the case of simple relaxation (to which Eq. (12) refers for $T^* = T$), the constancy of certain ratios is implied in Eq. (12)

$$\frac{Z(\epsilon'', T; T^*, t^*)}{Z(\epsilon', T; T^*, t^*)} = \frac{F(\epsilon'', T)}{F(\epsilon', T)}, \quad (13a)$$

$$\frac{Z(\epsilon, T''; T^*, t^*)}{Z(\epsilon, T'; T^*, t^*)} = \frac{F(\epsilon, T'')}{F(\epsilon, T')}, \quad (13b)$$

$$\frac{Z(\epsilon, T; T_2^*, t^*)}{Z(\epsilon, T; T_1^*, t^*)} = \frac{G(T_2^*, t^*)}{G(T_1^*, t^*)}, \quad (13c)$$

$$\frac{Z(\epsilon, T; T^*, t_2^*)}{Z(\epsilon, T; T^*, t_1^*)} = \frac{G(T^*, t_2^*)}{G(T^*, t_1^*)}. \quad (13d)$$

Again the constancy of these four ratios means geometrically the similarity of certain curves derived from the stress-temperature-time relationships.

The factorization expressed in Eq. (12) shows clearly that all isometrics, for various times t^* or temperatures T^* of pre-relaxation, differ only by a numerical factor $G(T^*, t^*)$. This factor is determined by the form of Boltzmann's memory function as explained in Section 4. Equation (12) also shows that the initial unstretched length should be taken for the computation of the extension ϵ . Thus the functional form of the equation of state may be obtained unambiguously from experiment.

The question of the *absolute value* of the stress is connected with the particular form of the memory function, i.e., with the statistical model used to explain the memory function.

Suppose the time elasticity of our material is given by a Maxwell body, i.e., a spring and a dashpot in series, leading to permanent set. Then one should extrapolate the time factor G to $t = 0$.

On the other hand, if the time elasticity is better described by a Maxwell body in parallel with a spring (leading to temporary set only), then one should extrapolate G to $t = \infty$. We restrict ourselves here to these two examples.

7. LIMITATIONS

The experimental results which yielded similarity for simple stress relaxation and for isometrics using pre-relaxation were obtained in air in the temperature range: 0–70°C.

The interesting work of Tobolsky and his associates⁹ at Princeton showed that relaxation and creep of rubber at temperatures above 100°C is caused by chemical changes involving the breaking and formation of primary valence bonds. The similarity law of Eq. (3) is valid. For our temperature range of 0–70°C, however, these processes are less important. In fact, in our experiments there was comparatively little permanent set left after taking the isometrics.

We wish to point out that there is always an observable decay of stress during the relaxation. Figures (6) and (7) of Tobolsky and Andrews¹⁰ are only in apparent contradiction to this statement. The apparent constancy of Z/T in these figures is caused by the small scale employed. Blown up to a larger scale, their results show also a stress decay.

Long duration creep and relaxation tests were carried out by Mooney, Wolstenholme, and Villars¹¹ at 35° and 70°C. For moderate extensions and duration their data are compatible with similarity.

It is planned to publish in greater detail in the near future the work summarized briefly in this paper. There, among other things, the role of the ordinary thermal expansion of rubber will be discussed.

⁹ Cf. A. V. Tobolsky and R. D. Andrews, *J. Chem. Phys.* **13**, 3 (1945).

¹⁰ See reference 9, pp. 9–10.

¹¹ M. Mooney, W. E. Wolstenholme, and D. S. Villars, *J. App. Phys.* **15**, 324 (1944).

The Theory of Permanent Set at Elevated Temperatures in Natural and Synthetic Rubber Vulcanizates

R. D. ANDREWS AND A. V. TOBOLSKY

Frick Chemical Laboratory, Princeton University, Princeton, New Jersey

AND

E. E. HANSON

Research Laboratory, The Firestone Tire and Rubber Company, Akron, Ohio

A molecular theory is developed to describe quantitatively the permanent set taking place in thin samples of vulcanized natural and synthetic rubbers held at constant extension at elevated temperatures. Permanent set is considered to be the result of the formation, through the action of molecular scission and cross-linking reactions, of a dual molecular network in the rubber sample, in which the network chains are of two types: chains which are at equilibrium when the sample is at its unstretched length, and chains which are at equilibrium when the sample is at its stretched length. According to the theory the amount of permanent set in a rubber sample is a function of only two quantities: the relative ratio of the number of chains of the two types, and the elongation at which the sample was held. Experimental data on permanent set for various rubber types and under different conditions are presented and are shown to be in good agreement with the theory.

CERTAIN materials, including natural and synthetic rubbers, after being deformed for finite periods of time and then released, do not return completely to their original dimensions. This phenomenon is called "permanent set."

There are many different types of measurements, made under a great variety of conditions, all of which are described as measurements of permanent set. The molecular causes of permanent set are not the same under all of these conditions. For this reason we shall restrict ourselves in the present article to a discussion of the permanent set taking place in thin rubber samples which are maintained at constant extension at elevated temperatures. We shall show that the permanent set occurring under these conditions is due to chemical changes which affect the structure of the rubber network and that a quantitative molecular description of permanent set can be accomplished by the use of concepts already developed concerning the nature of these deteriorative chemical processes.

MOLECULAR SCISSION AND CROSS-LINKING: THEIR MEASUREMENT AND RELATION TO PERMANENT SET

The changes in physical properties that occur in hydrocarbon rubbers at elevated temperatures are due to the simultaneous existence of chain scission and cross-linking reactions induced by

oxygen.¹ Depending upon whether chain scission or cross-linking is faster in each individual case, various rubber types soften or harden as a result of exposure to air at elevated temperatures. Butyl vulcanizates tend to soften; butadiene-styrene rubbers (GR-S, for example), butadiene-acrylonitrile rubbers, and Neoprene tend to harden; natural rubber (Hevea) tends first to soften, then to harden. Intermittent measurements of the modulus of a sample maintained in a relaxed state during exposure to the elevated temperature ("intermittent relaxation") provides a method of measuring the *net* rate of scission and cross-linking.

On the other hand it is possible to isolate the scission reaction by measuring the decay of stress in a sample maintained at constant extension at the elevated temperature ("continuous relaxation"). The decay of stress is a direct measure of the cutting of the molecular chains of the network. The cross-linking of chains, which is occurring simultaneously, tends to take place, to a first approximation, in such a way that it causes no increase of stress at constant extension. In other words, new network chains are being formed by a process of oxido-vulcanization which are in equilibrium when the sample is at the

¹ A. V. Tobolsky and R. D. Andrews, *J. Chem. Phys.* **13**, 3 (1945).

stretched length (by network chains we refer to the portions of chain molecules between contiguous network junctures).

The qualification that the cross-linking has no effect on the continuous relaxation curve "to a first approximation" is added because of the fact that in continuous relaxation experiments carried out for extended times an increase in stress is observed for some rubbers notably GR-S, Butaprene N, and Neoprene, following the initial almost complete decay of stress. This stress increase is reproducible with the same accuracy as the earlier part of the curve in which the stress decreases. A contraction of length has been observed to accompany embrittlement in unstretched thin strips resulting from heat aging. General theoretical considerations suggest that a volume contraction is probably taking place. Such a volume contraction would appear to be the most reasonable explanation of the observed stress increase, although this has not yet been investigated experimentally. It is apparent from the Neoprene-gum curve in Fig. 1 that the increase in stress may eventually become quite appreciable. In the case of Hevea gum the stress first decays completely to zero and later begins to increase as embrittlement of the stock sets in, following the initial softening. In Butyl stocks, which soften progressively, the stress decays completely to zero and no stress increase whatever is observed in the continuous relaxation curves even when they are carried out to very extended times. In the continuous relaxation curves of Hevea and Butyl stocks which appear later in this paper, no error should arise from this source, since in the time intervals to which the Hevea curves were carried out, the stock was still softening, and since the effect is not found at all in Butyl stocks. In the case of the GR-S tread stock, the relaxation curve was not carried out to times where this effect becomes significant. Nevertheless, it is well to mention this effect in any general statement concerning the nature of continuous relaxation curves.

Let us now consider the molecular changes taking place in a rubber sample which is being held at constant elongation at an elevated temperature, in terms of their relation to the phenomenon of permanent set. Immediately after extension the individual molecular chains are all

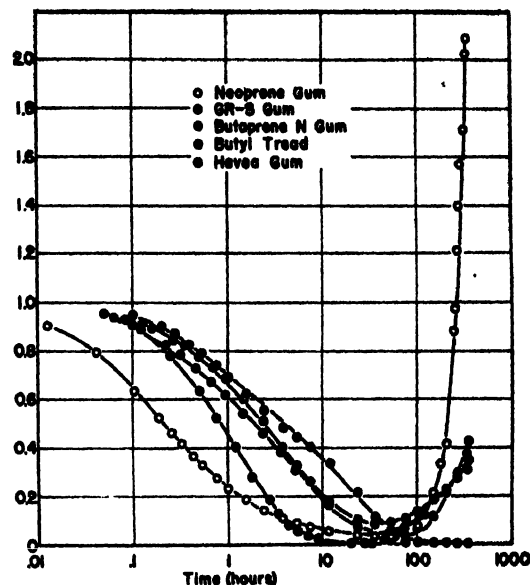


FIG. 1. Continuous relaxation curves for extended periods for various rubber types; 130°C; 50 percent elongation.

stretched and the equilibrium condition for all chains corresponds to the unstretched length of the sample. As the rubber remains at its constant elongation, cutting of the original chains takes place and the stress relaxes; new chains are formed by cross-linking which are at equilibrium when the rubber sample is at that particular elongation, and some of these newly formed chains are cut also. Thus at the beginning, the network chains in the rubber sample were all at equilibrium at the unstretched length; if the stress (due to chains being held out of equilibrium) eventually decays completely to zero, all the chains in the network are at equilibrium at the stretched length. At any intermediate stage between these two extremes, however, the molecular network in the rubber sample is a dual or "hybrid" network, containing chains of two types: (1) chains which are at equilibrium when the sample is at its unstretched length, (2) chains which are at equilibrium when the sample is at its stretched length. When the external constraint holding the sample at constant extension is removed, the chains of the first type tend to pull the sample back to its original unstretched length, whereas the chains of the second type tend to hold the sample out at its stretched length. As a result, the length which the sample actually assumes will be intermediate between its un-

stretched and stretched lengths and its value will be the length at which the stresses on each of the two sets of chains exactly balance each other.

DERIVATION OF PERMANENT SET EQUATION

Having formulated the theory of permanent set in the above qualitative way, it remains only to put the theory into mathematical form. This is done in the following way: For the case of a "homogeneous" network, in which all the network chains are at equilibrium at the same sample length, the equation relating stress and elongation in terms of the concentration of network chains in the rubber, is given as follows:

$$f = skT \left\{ \left(\frac{l}{l_u} \right)^2 - \frac{l_u}{l} \right\}, \quad (1)$$

where f is stress per unit *attained* cross-sectional area, s is the number of network chains per cc of the rubber, k is Boltzmann's constant, T is absolute temperature, and l and l_u are the attained length and unstretched length (the sample length at which the chains are at equilibrium) of the rubber sample, respectively. This equation is derived from the kinetic theory of rubber elasticity and is equally applicable both to extension and compression, i.e., l may be either greater or less than l_u . A positive value of f indicates a stretching force, and a negative value a compressing force. This equation should apply to both types of chains in the rubber network. Since the equation is derived from the statistics of a single chain, it is not necessary for either type of chain to form a continuous network within the rubber sample; it is thus perfectly applicable when one type of chain is predominant in the rubber network, chains of the other type occurring only here and there, in a relatively isolated way.

Let us denote the unstretched and extended lengths of the rubber sample by l_u and l_x , respectively, and the "set" length by l_s . We may then write for the two sets of chains in the rubber the two equations

$$f_u = s_u kT \left\{ \left(\frac{l_s}{l_u} \right)^2 - \frac{l_u}{l_s} \right\}, \quad (2)$$

$$f_x = s_x kT \left\{ \left(\frac{l_s}{l_x} \right)^2 - \frac{l_x}{l_s} \right\}, \quad (3)$$

where f_u is the stress per unit attained cross section on the chains which have their equilibrium position at the unstretched length l_u , and s_u is the number of those chains; f_x and s_x are defined correspondingly for the chains which are at equilibrium (or unstressed) at the extended length l_x . Now the condition for equilibrium in a sample with permanent set, as mentioned before, is that the retractive force exerted by the chains of type "u" is just balanced by the stretching force exerted by chains of type "x"; or, stated mathematically, that

$$f_u = -f_x. \quad (4)$$

Substituting in (4) the expressions for f_u and f_x from Eqs. (2) and (3), we obtain after simplification

$$\frac{s_u l_x^2}{s_x l_u^2} = \frac{l_x^3 - l_s^3}{l_s^3 - l_u^3}. \quad (5)$$

Permanent set is generally expressed as the percentage of the deformation retained by the sample; stated mathematically

$$\% \text{ Permanent Set} = \left(\frac{l_s - l_u}{l_x - l_u} \right) \times 100. \quad (6)$$

This may also be written in the form

$$\% \text{ Permanent Set} = \frac{(l_s/l_u - 1)}{(l_x/l_u - 1)} \times 100, \quad (7)$$

which is more satisfactory for the present purpose. It is to be noted that the denominator in (7) is a constant whose value depends on the elongation at which the sample is being held in the experiment. Thus only l_s/l_u must be evaluated to calculate permanent set. We can obtain l_s/l_u from Eq. (5) by algebraic transformation, giving the result

$$\frac{l_s}{l_u} = \left[\frac{(l_x/l_u)^2 - 1}{(s_u/s_x)(l_x/l_u)^2 + 1} + 1 \right]^{1/3}. \quad (8)$$

This expression for l_s/l_u may now be substituted in Eq. (7), giving the following equation for permanent set

$$\% \text{ Permanent Set} = \left\{ \left[\frac{C_1}{(s_u/s_x)C_2 + 1} + 1 \right]^{1/3} - 1 \right\} C_2, \quad (9)$$

where C_1 , C_2 , and C_3 are constants whose values depend on the elongation of the sample

$$C_1 = \left(\frac{l_x}{l_u}\right)^3 - 1; \quad C_2 = \left(\frac{l_x}{l_u}\right)^2; \quad C_3 = \frac{100}{(l_x/l_u) - 1}.$$

Thus it is to be noted that according to this theory the amount of permanent set in a rubber sample depends upon the elongation of the sample as well as upon the ratio of s_u to s_x .

In order to verify this relation experimentally it is necessary to evaluate the ratio s_u/s_x in terms of continuous and intermittent relaxation curves. This can be done very simply and directly: the value of s_u/s_x at any time t is obtained by dividing the stress value given by the continuous curve at that time by the difference between the stress values of the continuous and intermittent curves. The reason for this is that the stress value given by the continuous curve is proportional to s_u , and that given by the intermittent curve is proportional to $(s_u + s_x)$, by the same factor or proportionality. There are actually no chains of type x in the intermittent sample, of course, but the number of chains of type u is equal to the sum of the number of chains of types u and x in the continuous sample. (The assumption made here, of course, is that the rates of scission and cross-linking are exactly the same in stretched and unstretched samples, so that at any given time the *total number* of chains is the same in both the intermittent and continuous samples.)

The stress data given in this paper are plotted as stress divided by initial stress, f/f_0 , rather than

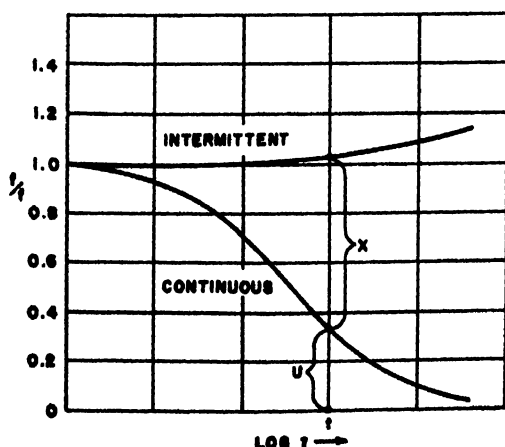


FIG. 2. Evaluation of U and X from continuous and intermittent relaxation curves.

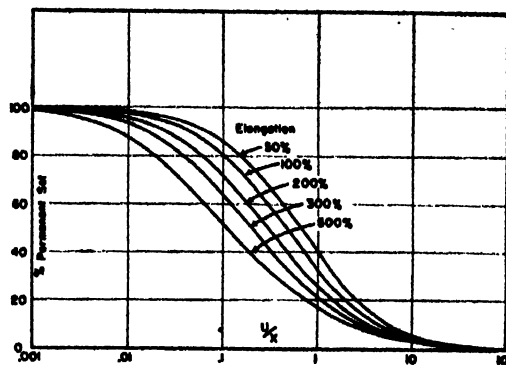


FIG. 3. Permanent set function for various elongations.

as absolute stress. The procedure for evaluating s_u/s_x is not affected by this fact, however. The f/f_0 value given by the continuous curve is designated as U , and the difference in the f/f_0 values of the intermittent and continuous curves as X . This is illustrated in Fig. 2. Thus, according to this notation

$$s_u/s_x = U/X, \quad (10)$$

and U/X may be substituted for s_u/s_x in Eq. (9), giving the following final equation for permanent set in terms of relaxation data

% Permanent Set

$$= \left\{ \left[\frac{C_1}{(U/X)C_2 + 1} + 1 \right]^{\frac{1}{3}} - 1 \right\} C_3, \quad (11)$$

where the constants have the same values as before. A graph of this permanent set function for a number of different elongations, plotted in the form of percent permanent set *versus* the logarithm of U/X , is shown in Fig. 3.

It might be well to add a few additional remarks at this point concerning the central assumption in the theory of permanent set presented here, i.e., that the cross-linking takes place in such a manner as to form chains which are relaxed when the sample is at the length at which the cross-links were formed. It is not possible to describe satisfactorily a single local cross-linking process mathematically; the effect must rather be considered as a statistical one, and it is necessary to say that the *average effect* of the cross-linking is to form new chains which are relaxed, just as is the case with the cross-linking taking place during vulcanization. Thus there are probably chains under a whole distribution of stress states in a

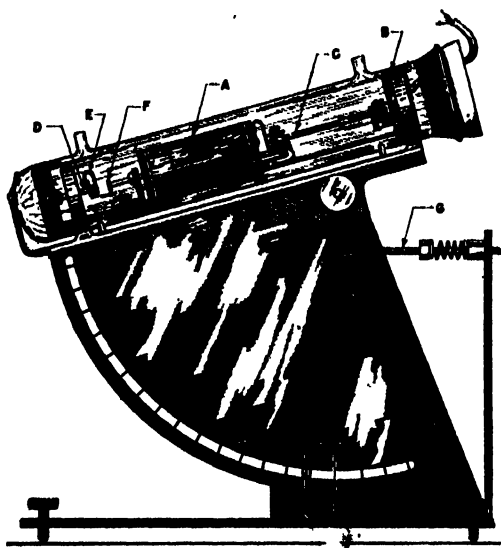


FIG. 4. Tilting relaxometer.

sample held at constant elongation for a period of time at high temperatures. This situation is considered to be equivalent, however, to the case of a network in which there are chains of only two distinct types—chains which are at equilibrium at the original unstretched length of the sample, and chains which are at equilibrium at the stretched length. This model makes possible a mathematically simple calculation of permanent set, just as the same sort of simplification makes possible the calculation of the theoretical stress-strain Eq. (1), obtained from the kinetic theory of rubber elasticity. Therefore, although the picture of a "dual" network employed in the derivation above undoubtedly does not describe exactly the character of the individual chains in the network, it should describe the statistical average behavior of the individual chains.

APPARATUS AND EXPERIMENTAL PROCEDURE

The continuous and intermittent relaxation curves included in this paper were obtained in part on a beam balance type of relaxation apparatus, and in part on an inclined plane type of relaxation apparatus. Duplicate relaxation curves from the two types of apparatus have been found to agree satisfactorily. A description of the beam balance apparatus and its operation has been given in a previous publication.²

² A. V. Tobolsky, I. B. Prettyman, and J. H. Dillon, *J. App. Phys.* 15, 380 (1944).

A sketch of the tilting apparatus is shown in Fig. 4. The apparatus consists of two parts: part A is a brass cart for carrying as many as five 250 g lead weights, and part B is a brass frame which surrounds the weight cart, but is separate from it. The frame and the cart fit inside a glass tube along which the cart is free to roll on its four wheels. The rubber samples (C) used in this apparatus are flat rings ($1\frac{1}{2}$ in. O.D., $1\frac{1}{2}$ in. I.D.), died from thin cured sheets (approximately .030 in. thick). Thin samples are used to insure complete penetration of oxygen at elevated temperatures, so that the scission and cross-linking reactions occur homogeneously throughout the samples. These samples are looped around two stainless steel pulleys, one attached to the upper end of the frame and the other to the upper end of the cart, as shown. Projecting from the lower end of the frame is a support post (D) with a phosphor bronze spring contact (E) at its upper end. The motion of the cart is arrested when a small wooden knob (F) fixed to the lower end of the cart touches this support post; this knob also closes the spring contact in touching the post. The support post is replaceable; posts of lengths to give the sample any elongations desired in the range of 0 percent to 200 percent are easily constructed.

The glass tube containing the apparatus is attached by clips to a tilting stand in the manner illustrated in Fig. 4. The apparatus is tilted by a crank shaft (G) which extends through the side of the temperature box and is connected with the tilting stand through a worm gear. The entire apparatus, tube and tilting stand, is installed in an air oven with a double glass door. The temperature is thermostatically controlled to 0.3°C , and air circulation is maintained by three small air blowers. The temperature is regulated with reference to a small iron-constantan thermocouple (not shown in sketch) which is attached to the side of the sample with fine thread. The measurement of the test temperature in this way, exactly at the position of the sample, was found to be necessary in order to obtain the most accurate and reliable results.

In carrying out a relaxation experiment the apparatus was first heated thoroughly at the test temperature. The tube was then removed from the oven, the sample (with thermocouple at-

tached) looped over the pulleys, the tube replaced in the oven with the stand at 0° tilt, the electrical leads from the spring contact connected to a neon glow bulb on the outside of the oven, and the oven closed again. When the thermocouple indicated that the sample had reached the test temperature, the sample was extended, and stress readings begun. In taking a stress reading, the angle of tilt was slowly decreased, tapping rapidly on the crank shaft at the same time, (to give the apparatus a slight vibration), until a suitable flicker was observed in the neon signal bulb (a flicker equally on and off was taken as the indication of balance.) The time was noted and the angle of tilt read by means of an indicator on the base of the tilting stand; the scale on the stand is marked in half degree units, and the angle of tilt was read to 0.1 degree. In continuous relaxation experiments (in which the sample is maintained continuously at the specified elongation) the angle of tilt was maintained between readings at some value greater than that necessary to balance the stress in the sample, and was never decreased below the angle of balance, so that there was no opportunity for the sample to contract to an elongation lower than that determined by the support post. In intermittent relaxation experiments (in which the sample remains unstretched at all times, except when a stress reading is taken), the tilt was set at zero degrees between readings, the apparatus being returned to the level position immediately after each stress reading. The time required to take a stress reading on an intermittent sample was generally 15 to 30 seconds.

The weight on the sample is equal to the total weight of the cart (cart plus slotted weights) multiplied by $\sin \alpha$, the sine of the angle of tilt. When it is desired only to calculate f/f_0 , the stress at time t divided by the initial stress, as a function of time (as was the case in the present work), this can be directly evaluated from the equation:

$$f/f_0 = \sin \alpha / \sin \alpha_0, \quad (12)$$

where α and α_0 are the angles of balance at times t and zero, respectively. Zero time was taken as the time of the first stress measurement, for both continuous and intermittent experiments.

Measurements of permanent set were made by



FIG. 5. Permanent set frame.

use of the brass frame shown in Fig. 5. The frame consists simply of square bars at top and bottom held by support rods at the side, and provided with bolt clamps to hold the samples. An idea as to the size of the frame can be gained from the fact that the distance between the inside edges of the top and bottom bars is $6\frac{1}{2}$ in. (Fig. 5 is drawn to scale.) The frame will hold as many as five samples.

The samples used were small strips $\frac{1}{4}$ in. wide and approximately 6 in. long, cut from the thin vulcanized sheets from which relaxation samples are obtained. Two fine silver ink marks were placed on the edge of each sample an appropriate distance apart, and the distance between marks was accurately measured for each sample with a millimeter rule. The samples were then fastened to the frame, stretched to the desired elongation, and the distance between marks again accurately measured. The frame was inserted in an air oven whose temperature was thermostatically controlled to $\pm 0.3^\circ\text{C}$, the temperature being indicated by a thermometer with its bulb adjacent to the frame. The frame was suspended from a nail on the back wall of the oven, by means of a small L-bracket attached at the end of the top cross-bar of the frame, so that the frame projected sidewise perpendicularly from the oven wall. When it was desired to remove a sample, at

various time intervals, the frame was removed from the oven, the sample quickly cut loose with scissors, and the frame immediately replaced in the oven. After the samples had cooled for 5 min., the distance between ink marks was again measured, and the permanent set of the sample calculated by the equation

$$\% \text{ Permanent Set} = \frac{l_s - l_u}{l_s - l_u} \times 100. \quad (13)$$

The original distance between marks l_u , the distance when the sample was stretched on the frame l_s , and the distance after removal of the sample from the oven l_u , were each measured to an accuracy of 0.1 mm and the permanent set was calculated to 0.1 percent. Appropriate time corrections were applied for the time required for the oven to reach temperature after inserting the frame and for the time in which the frame was out of the oven to remove samples; the corrections necessary were small in each case.

In one case, i.e., the permanent set measurements on GR-S tread stock at 130°C, a slightly different procedure was followed in cutting the samples from the frame. In these measurements the samples were cut loose at one end without opening the oven, by the use of a razor blade attached to a rod which was manipulated through a small aperture in the back of the oven. One minute later the oven was opened and the sample was cut free from the frame with scissors, without removing the frame from the oven. The permanent set was measured 5 min. after removal of the sample from the oven, as before, and in all other respects the procedure was the same as that used for the other stocks. This modified procedure was used in order to avoid any cooling of the sample before cutting, which would result in the formation of "secondary" interchain bonds in

the sample, which in turn would cause an error in the permanent set value. This effect will be discussed in more detail in the section on experimental results.

The time interval of 5 min. between removal of the sample from the oven and measurement of permanent set was chosen to allow the sample to cool completely to room temperature. The recovery observed on standing of the samples at room temperature was quite small, as can be seen from the data in Table I. Recovery was measured at intervals of one hour, one day, and one week after the original permanent set measurement. Samples of the four different rubber stocks studied in this paper were used, each of which was exposed for two hours at 130°C, at 50 percent elongation. It is seen that the recovery was less than 3 percent in every case after a week had elapsed, being particularly small in the case of Butyl gum.

EXPERIMENTAL RESULTS

Results of permanent set measurements, in which experimental permanent set values are compared with permanent set predicted from continuous and intermittent relaxation curves by means of the permanent set theory developed above, are shown in Figs. 6–10. These experiments were planned to include as many variables as possible: different polymer types, gum and tread type stocks, different temperatures, and different elongations. Four different rubber stocks were used in these experiments: Hevea gum, Butyl gum, Butyl tread, and GR-S tread. The compounding and cure of these stocks is given in Table II.

Permanent set results for Hevea gum stock at 130°C and 50 percent elongation are shown in Fig. 6. The continuous and intermittent relaxation curves are drawn in as solid lines; and the theoretically predicted curve of permanent set as a function of time, calculated from the relaxation curves, is drawn as a dashed line. Experimental permanent set points are shown as circles, each point, of course, being obtained from a separate sample. It is seen that the agreement between experiment and theory is excellent.

Similar data for Butyl gum and Butyl tread at 130°C and 50 percent elongation are shown in Figs. 7 and 8. Although the agreement between

TABLE I. Recovery at room temperature.

Stock	Percent permanent set after recovery times of:			
	0 hr.	1 hr.	1 day (22 hr.)	1 week (169 hr.)
Hevea gum	58.3%	57.9%	56.6%	56.0%
Butyl gum	42.9%	42.9%	42.5%	42.3%
Butyl tread	57.3%	57.3%	56.3%	55.5%
GR-S tread	60.9%	60.3%	59.3%	58.9%

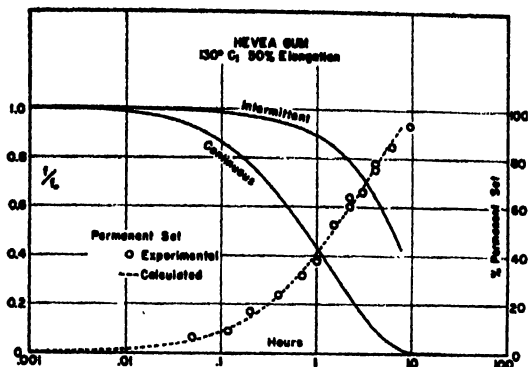


FIG. 6. Permanent set as a function of time for hevea gum stock at 130°C and 50 percent elongation.

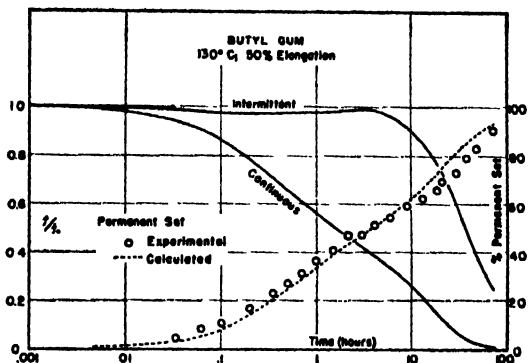


FIG. 7. Permanent set as a function of time for butyl gum at 130°C and 50 percent elongation.

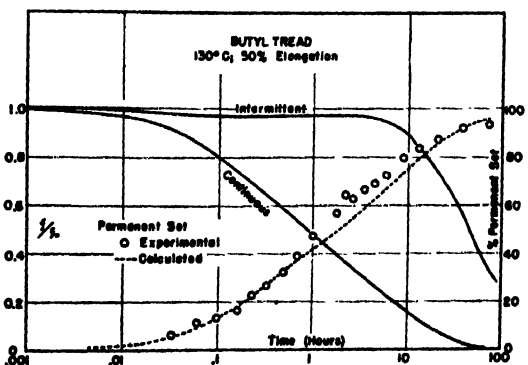


FIG. 8. Permanent set as a function of time for butyl tread; 130°C; 50 percent elongation.

the experimental points and the theoretically calculated permanent set curves is not as precise here as was the case for Hevea gum, the agreement must nevertheless be regarded as satisfactory. Duplicate permanent set points were obtained for Butyl gum at 1.00 hr., 2.20 hr., and 4.10 hr. The agreement between these duplicate samples was so close that it was impossible to

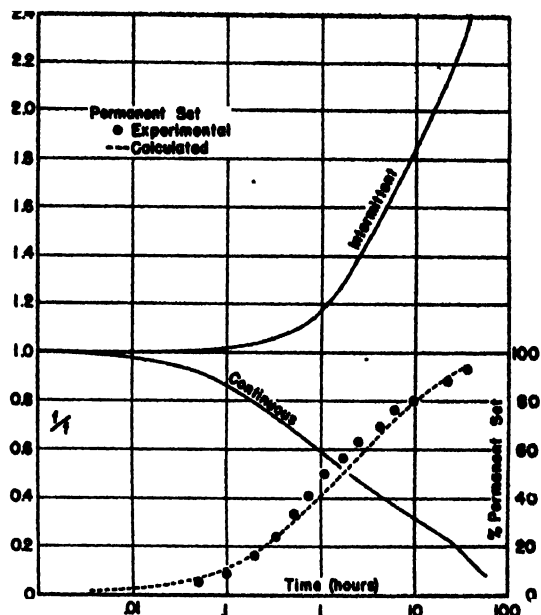


FIG. 9. Permanent set as a function of time for GR-S tread; 130°C; 50 percent elongation.

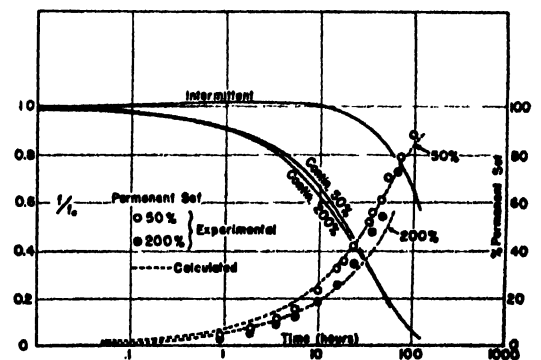


FIG. 10. The effect of elongation on permanent set. Hevea gum at 50 percent and 200 percent elongation; 100°C.

draw two separate points; for this reason these points are indicated by half circles in the graph.

Relaxation and permanent set data are shown for GR-S tread at 130°C and 50 percent elongation in Fig. 9. In this stock molecular cross-linking takes place more rapidly than scission, unlike the Hevea and Butyl stocks, in which the reverse is true. Nevertheless, it is seen that the agreement between experimental and calculated permanent set values is rather good in this case also.

It will be remembered that in this particular experiment the permanent set samples were cut at the test temperature while the oven remained

closed, rather than immediately after removal of the set frame from the oven, as was the case in the other permanent set measurements reported here. This was found to be necessary because of the formation of stable "secondary bonds" in the

TABLE II. Compounding of stocks.

Type stock	Hevea gum	Butyl gum	Butyl tread	GR S tread
Polymer:				
Hevea smoked sheet	100			
GR-I		100	100	
GR-S				100
Sulfur	3	2	2	2
Zinc oxide	3	5	5	5
Stearic acid	0.5	1.5	3	3
Captax	1	0.5	0.5	1.5
Tuads		1	1	
Agerite powder	1			
Bardol				5
Channel black			50	50
Cure	75 min. at 270°F (132.2°C)	75 min. at 300°F (148.9°C)	100 min. at 300°F (148.9°C)	100 min. at 270°F (132.2°C)

sample as a result of the cooling which took place on removal of the frame from the oven, before the samples could be cut. These so-called secondary bonds are not primary chemical valence bonds, but rather are weak physical inter-chain bonds (possibly of a dipole-dipole or van der Waals type) which are quite unstable at very high temperatures, but whose stability increases as the temperature is lowered. Since in terms of stress phenomena (such as permanent set) a secondary bond which has not broken has the same effect as a primary valence cross-link, the formation of stable secondary bonds by cooling of a sample at its extended length will be the same as if an increased amount of primary valence cross-linking had taken place; i.e., the observed permanent set will be increased. It is found experimentally that permanent set values obtained from GR-S tread samples cut from the frame immediately after removal from the oven are a few percent greater than the values obtained from samples cut in the oven at test temperature; and when the frame is removed from the oven and the samples allowed to cool a few minutes before cutting, the permanent set values obtained are very markedly greater than the values obtained from duplicate samples

exposed for the same length of time to the elevated temperature but cut at test temperature with the oven closed. GR-S tread stock shows large secondary bond effects in general. When permanent set samples of the other stocks were cut at the test temperature in the same way, the permanent set values obtained from Hevea gum and Butyl gum samples did not differ significantly from the results in Figs. 6 and 7; permanent set values of Butyl tread samples were slightly less than the values in Fig. 8, thus slightly improving the agreement between experiment and theory. There was apparently a slight secondary bond error in the case of Butyl tread; in general it is observed that secondary bond effects are greater in tread stocks than in the corresponding gum stocks. It would undoubtedly be best to cut permanent set samples at test temperature as a standard procedure, to avoid any error due to this source.

In Fig. 10, permanent set data are shown for Hevea gum at 100°C at both 50 percent and 200 percent elongation. Theoretical permanent set curves have been calculated for both elongations from the continuous relaxation curves at the two elongations and the intermittent curve (obtained at 50 percent elongation) which are shown in the graph. The same intermittent curve was used for both calculations since the elongation of the intermittent sample is really 0 percent in both cases and intermittent curves in which the stress is measured at different elongations should give the same f/f_0 curve. It is seen that the experimental permanent set points for 50 percent elongation agree well with the calculated 50 percent curve. The 200 percent elongation experimental points agree satisfactorily with the calculated 200 percent curve at the earlier times, but for some reason rise somewhat above the calculated curve at the later times, having values more nearly like those of the 50 percent points than is predicted from the relaxation curves. The 200 percent experimental points appear to give a smooth curve, just as the 50 percent points do, so that apparently the effect is not fortuitous. The explanation for this deviation is not known. There is a definite difference between the 50 percent and 200 percent experimental permanent set points, however, the 200 percent points being

lower than the 50 percent points, which agrees with the predictions of the theory.

The generally satisfactory agreement of experimental permanent set data with predicted permanent set, calculated from continuous and intermittent stress relaxation curves by means of the permanent set theory developed in this paper, appears to offer a rather good confirmation of the general correctness of the molecular picture of permanent set at elevated temperatures upon which this quantitative permanent set theory is based.

ACKNOWLEDGMENT

This work was done at Princeton University with the cooperation of the Firestone Physics Research Group. The authors wish to thank F. S. Grover for his work in constructing the tilting relaxometer. Valuable suggestions made by Professor H. S. Taylor and Dr. J. H. Dillon are gratefully acknowledged. The permission of the Firestone Tire and Rubber Company, who sponsored this work, to publish these results is appreciated.

Calendar of Meetings

May

- 6-7 Inter-Society Color Council, New York, New York
- 6-10 Society of Motion Picture Engineers, New York, New York
- 10-11 Acoustical Society of America, New York, New York
- 27-29 American Geophysical Union, Washington, D. C.

June

- 2-5 American Society of Refrigerating Engineers, St. Paul, Minnesota
- 10-14 American Society for X-Ray and Electron Diffraction, Silver Bay (Lake George), New York
- 13-14 Institute of Aeronautical Sciences, Detroit, Michigan
- 17-20 American Society of Mechanical Engineers, Detroit, Michigan
- 20-22 American Association of Physics Teachers, St. Louis, Missouri
- 20-22 American Physical Society, Chicago, Illinois
- 20-23 Society for the Promotion of Engineering Education, St. Louis, Missouri

- 24-28 American Institute of Electrical Engineers, Detroit, Michigan
- 24-28 American Society for Testing Materials, Buffalo, New York

July

- 12-13 American Physical Society, Berkeley, California
- 17-19 American Society of Civil Engineers, Spokane, Washington

August

- 14-15 Institute of Aeronautical Sciences, Los Angeles, California
- 19-20 Mathematical Association of America, Ithaca, New York

September

- 9-13 American Chemical Society, Chicago, Illinois
- 16-20 The Instrument Society of America, Pittsburgh, Pennsylvania
- 18-21 Illuminating Engineering Society, Quebec, Canada
- 19-21 American Physical Society, New York, New York
- 30-October 3 American Society of Mechanical Engineers, Boston, Massachusetts

Crystallization of Unvulcanized Rubber at Different Temperatures

LAWRENCE A. WOOD AND NORMAN BEKKEDAH.
National Bureau of Standards, Washington, D. C.

The crystallization and melting of unvulcanized natural rubber in the unstretched state have been investigated at different temperatures. Change of volume has been used as a quantitative measure of the extent of crystallization, and mercury-filled dilatometers containing the rubber have been used for the volume measurements. Crystallization was observed to occur at temperatures between -50° and $+15^{\circ}\text{C}$ and to be most rapid at about -25°C . The final decrease of volume on crystallization was usually found to lie between 2.0 and 2.7 percent. The melting of the crystalline rubber was found to occur over a range of temperature and to be strongly dependent on the temperature at which the crystals were formed. The temperature at which the beginning of melting occurs is from 4° to 7° above the temperature of crystallization. The range of melting is about 35° at the lowest temperatures and decreases to about 10° at the highest. The same range of temperature of melting is obtained regardless of the extent of the crystallization.

I. INTRODUCTION

CRYSTALS may be formed in natural rubber under quite varied experimental conditions. Different combinations of stretching and cooling have been used to induce crystallization in unvulcanized and in vulcanized rubber. The appearance and disappearance of crystals have been studied by observations of the volume,^{5, 8} heat capacity,⁹ light absorption,¹³ birefringence,^{40, 58, 63, 67, 74} x-ray diffraction,²² and hardness⁵³ and other mechanical properties.⁴² There has, however, been no comprehensive study of the effect of temperature on the crystallization. The present investigation was undertaken in order to explore this field.

In the work reported here it has been the aim to study the crystallization at different temperatures under the simplest possible conditions. The main features of the crystallization of vulcanized rubber have been shown⁸ to be similar to those of the crystallization of unvulcanized rubber, the vulcanization causing a rise in the rate of crystallization. Consequently unvulcanized rubber was selected for study. Stretching obviously adds complication to the experimental conditions, and so was not employed. Of the different methods of measuring crystallization, it seems that change of volume is the simplest and best adapted to yielding quantitative data about the course of the crystallization or fusion. The present work therefore is concerned with a general study of the volume changes in unvulcanized rubber at different temperatures.

II. EXPERIMENTAL PROCEDURE

The volume changes of the specimens were measured by the use of dilatometers with confining liquids. The technique incorporates some improvements over that previously used.^{5, 8} The dilatometers were made of Pyrex glass and each consisted merely of a bulb with a capillary tube sealed to it. Changes in volume could be calculated from measurements of changes in the height of the confining liquid in the capillary tube. Mercury was selected as the confining liquid because it seems to have no effect on rubber, even when the two are in contact for periods of several years. Some absorption, swelling, and softening usually occur with other liquids. The specimens were strips cut from plantation smoked sheet rubber.

The construction of the dilatometers and the procedure of filling them were somewhat improved during the course of the work reported here. In the later dilatometers the capillary tubes were about 2 mm in inside diameter and about 500 mm in length. They were specially selected capillaries, but were not of the precision-bore type recently developed. They were graduated in millimeters along their length. The tubes were calibrated by sealing a stopcock to one end, and observing the lengths of weighed amounts of mercury at five or more different points along each tube. The maximum variation in cross-sectional area along a single capillary was found to be about 3 percent in most cases. Each capillary was then sealed to a glass tube about 15 mm

in inside diameter, the other end of the tube being left open. Into the glass tube were inserted a weighed rubber specimen and a hollow glass bulb about 40 or 50 mm long. Finally the open end of the glass tube was sealed off to form the bulb of the dilatometer.

This construction made it possible to avoid heating of the specimen during the sealing operation, without unduly increasing the net volume of the dilatometer bulb. In most cases the volume of the specimen was between two and three times the volume of the mercury in the bulb of the dilatometer. The effect of the insertion of the hollow bulb is equivalent to a mere reduction of the net volume of the dilatometer bulb. Since the inner hollow bulb was made of the same type of glass as the rest of the dilatometer, neither its volume nor its expansivity entered directly into any of the calculations.

The dilatometer was next evacuated for several days to remove gas from the specimen. Mercury was then admitted to the evacuated system through a two-way stopcock, until it filled the dilatometer bulb and stood at a suitable height in the capillary. Occasionally it was necessary to carry out further removal of gas until the height of the mercury in the capillary did not change by more than a few millimeters as the pressure was again raised to that of the atmosphere. Calculation showed that the gas remaining under these conditions would have a negligible effect on the observed specific volume and expansivity. The dilatometer was weighed before and after the addition of the mercury.

The volumes of specimen and mercury at 25°C were determined from their weights and densities. The densities of the rubber specimens at this temperature were measured by the method of hydrostatic weightings. The volume of the dilatometer up to the level of the mercury in the capillary at this temperature was obtained as the sum of the volumes of the mercury and the specimen. At any other temperature the volume up to the same point on the capillary was found from the known expansivity of Pyrex glass. The volume of the dilatometer up to any other point on the capillary could then be calculated from the calibration of the capillary. The volume of the specimen corresponding to each observation was obtained by subtracting from this volume, the

volume of the mercury at the temperature of the observation.

After the specimens and dilatometers had been prepared, they were usually placed in a stirred bath of about 6 liters of alcohol cooled by solid carbon dioxide to a temperature of about -39°C , an approximate lower limit set by the freezing of the mercury. Observations of the height of the mercury in the capillary were made as the temperature was raised to $+55^{\circ}\text{C}$ at a rate of the order of 0.5°C per minute. At the lower temperatures the bath was warmed at about this rate by the heat from its surroundings. Beginning at about -10°C additional heat was furnished by means of a knife-type immersion heater controlled by a Variac. The alcohol was usually replaced with water for measurements above room temperature. In this manner the volume-temperature relation of the amorphous rubber was obtained. This relation, which is linear over a considerable range of temperature, will be described and discussed in greater detail later in connection with the melting of crystalline rubber.

It was found in preliminary experiments that at a heating rate of 0.5°C per minute or less no observable differences from the equilibrium values of the volume could be noted. At a heating rate of about 1°C per minute the apparent volume lagged behind the equilibrium value by an amount equivalent to about 1°C . At a heating rate of about 2° per minute, the lag was about 2°C .

The dilatometers were next placed in a small unstirred alcohol bath in a refrigerator at the temperature at which it was desired to crystallize the rubber. Observations of the height of the liquid in the capillary tube were made periodically and usually continued until no further volume change could be noted.

When it was observed in this manner that the rate of crystallization had become inappreciable, the specimens were placed in the stirred bath of alcohol cooled by solid carbon dioxide to a temperature of about -39°C . Observations of the volume were made as before, as the temperature was raised, and the volume-temperature relation of the crystalline rubber was obtained.

In a number of cases, to be described in detail later, there were deviations from this standard procedure for the purpose of studying the effect

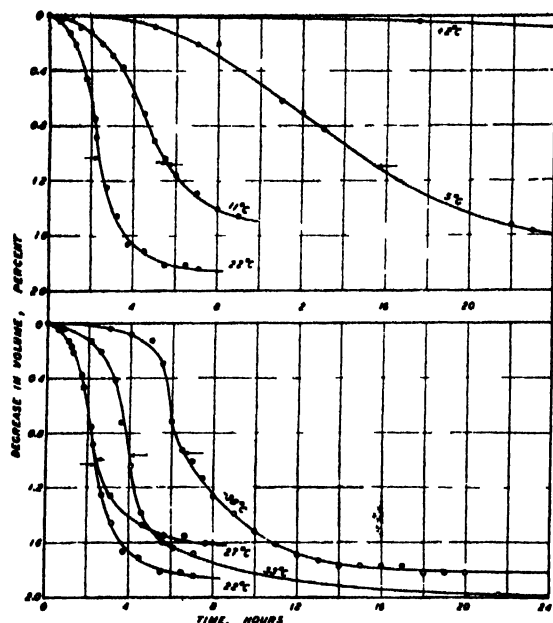


FIG. 1 Crystallization of rubber at different temperatures as indicated by decrease in volume. The arrows indicate the estimated values for half the total decrease of volume at each temperature.

of varying some of the experimental conditions, or because the importance of some of the exact conditions was not realized at the time the experiment was performed.

III. EXPERIMENTAL RESULTS

1. Rate of Crystallization

When crystallization was carried out according to the procedure just described it was found that the rate was very low at first, increased to a maximum, and decreased to a negligible value. The volume-time relation at constant temperature thus has the sigmoid shape which has previously been given^{4, 6, 69} for crystallization near 0°C. Figure 1 shows two families of such curves, one for crystallizing temperatures of -22°C and above and the other for -22°C and below.

Since the volume is changing very slowly in the final stages, a determination of the time required for completion is somewhat uncertain. The time required for one-half the total volume change is much more precisely determinable since the curves are usually steepest near this point. The time required for half the volume change has been measured at a number of different temper-

atures. The reciprocal of this time, it can be readily seen, is a measure of the average rate of crystallization during the first half of the process. It is plotted as a function of temperature in Fig. 2.

2. The Melting of Crystallized Rubber

When the volume-temperature relation of crystalline rubber is investigated, as already described, curves of the type shown in Fig. 3 are obtained for increasing temperatures.

It can readily be seen that melting of the crystals occurs over a range of temperature, and that the range is markedly dependent on the temperature at which crystallization has occurred.⁷ The temperature corresponding to the beginning of melting is that at which the volume begins to deviate from the curve representing the normal thermal expansion of crystalline rubber. Over the range of temperatures under consideration here the normal expansion curve is linear, and the temperature at which deviation begins to occur can be located within about one-half a degree. Similarly melting is complete when the volume again reaches the nearly linear curve which represents the volume-temperature relation of amorphous rubber and which was evaluated earlier.

There was found to be no evidence of any time lag or delay in melting of the crystals after the establishment of thermal equilibrium. For example, at any constant temperature in the melting range the volume was never observed to increase with time. In this range if the temperature is lowered, the volume-temperature relation is observed to be linear and intermediate between the amorphous and crystalline volumes. Rubber

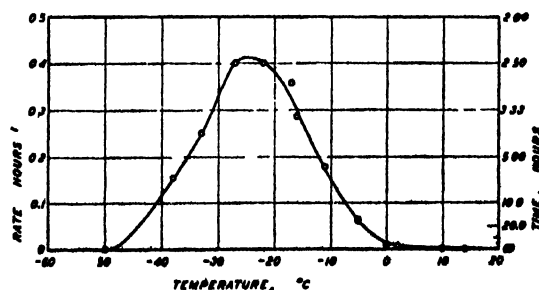


FIG. 2. Rate of crystallization of rubber. The rate plotted is the reciprocal of the time required for one-half the total volume change.

in the partially-melted state, however, can undergo additional crystallization with a resulting decrease of volume. If the temperature at which the additional crystallization occurs is the same as that at which the first crystallization has occurred, the melting curve is identical with that obtained from the complete melting of the first crystals. If the additional crystallization occurs at a different temperature from the first crystallization, two different ranges of melting will be observed, corresponding independently to the two temperatures of crystallization. Further discussion of this effect will be given below in the section on stark rubber and is illustrated in Fig. 7 in that section.

The additional crystallization possible in partially-melted rubber mentioned in the preceding paragraph is considered in more detail in the section dealing with recrystallization. At this stage, however, it should be pointed out that the phenomenon may give rise to difficulty in determining the upper end of the melting range itself.

The rate of crystallization at certain temperatures was sufficiently great to permit significant recrystallization during the time that the temperature was being raised through the melting range of the original crystals. Since the recrystallization took place at a higher temperature than the original crystallization, the new crystals melted at a higher temperature than the original crystals, and the apparent upper end of the melting range was displaced upward by an amount which depended on the rate of temperature rise during melting.

When the initial crystallization occurred below 0°C , therefore, the upper end of the melting range was determined from a series of experiments at each temperature of crystallization. A stirred bath was adjusted to the approximate temperature of the upper end of the melting range. The dilatometer was then removed from the bath in which the original crystallization had been taking place and plunged into the stirred bath. The volume of the specimen increased because of both thermal expansion and the melting of the crystals. In less than five minutes, a reasonable time for the establishment of temperature equilibrium, the volume became constant. After an interval of time, the volume began to decrease

because of recrystallization at the temperature of the stirred bath.

If the value of the volume when temperature equilibrium had become established and before observable recrystallization had occurred was the same as that originally observed for amorphous

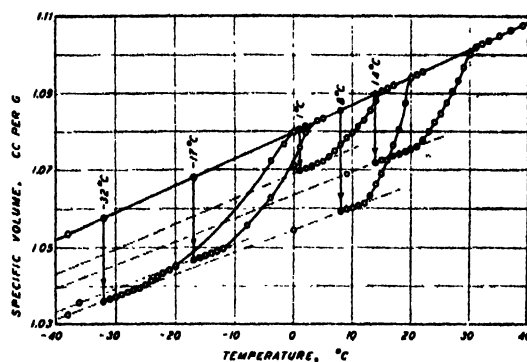


FIG. 3. Crystallization of rubber at different temperatures and subsequent melting. The change of specific volume on crystallization is indicated by the vertical lines joining the dark circles.

rubber at the corresponding temperature, it was concluded that the temperature of the stirred bath was at or above the upper end of the melting range. If the volume at equilibrium was less than that of the amorphous rubber at that temperature the conclusion was drawn that the melting was incomplete and that the temperature of the bath was below the upper end of the melting range. As might be expected the volume at equilibrium was never found to be greater than the volume of the amorphous rubber. A series of experiments of this sort served to locate the upper end of the melting range corresponding to a single temperature of original crystallization, and an additional series was required for each other temperature.

If the initial crystallization occurred at 0°C or above, the rate of recrystallization during melting was found to be so small as not to be significant during the time the specimen was in the melting range when the usual heating rates were employed. At these temperatures the range was of the order of 10° or less, and the time required to cover the range was usually between 20 and 30 minutes. The upper end of the melting range was determined from curves of the type shown in Fig. 3 for all cases where the initial crystallization was at 0°C or above.

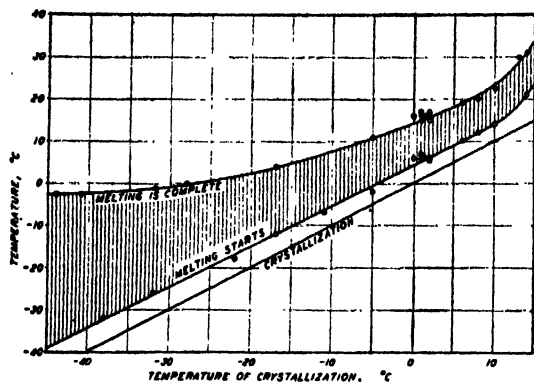


FIG. 4. Melting range of crystalline rubber as a function of the temperature of crystallization. The lower line shows the crystallization temperature also as ordinate for easier comparison with the range of melting.

The results of a number of individual experiments of this sort are plotted in Fig. 4 which shows the melting range as a function of the temperature of crystallization.

In one series of experiments a specimen was successively crystallized and melted 16 times. After each melting the temperature was raised by different amounts varying from 1° to 40° above the end of the melting range. In all cases the same melting range was obtained on subsequent crystallization at a fixed temperature.

3. Partial Crystallization

The melting range of partially-crystallized rubber has been found to be the same as that obtained when the crystallization is more nearly complete. Data regarding this point were obtained from a series of experiments in which the specimen was kept at 2°C for different lengths of time, and the melting range of the resulting crystals then measured. In the first experiment after 20 days of crystallization the volume decrease amounted to 2.7 percent. In the second experiment crystallization was interrupted at the end of 9 days when the volume change amounted to only 1.0 percent (or 37 percent of the volume change in the first experiment). In the third experiment by interruption of the crystallization after 1 day the volume change was limited to 0.1 percent (or 3.1 percent of that in the first experiment). The melting range in all three cases was 6° to 16°C within the accuracy of determination. On the basis of these experiments the crystallization in some cases was not carried

completely to the point where the volume had ceased to change at a significant rate. Cases of this sort will be observed in Fig. 3.

4. The Recrystallization of Rubber

It will be noted that in the cases previously described the specimens were heated to at least 55°C immediately before crystallization. The purpose of this heating was to destroy centers of crystallization which might have been persisting from previous crystallizations, and which would have altered the rate of crystallization. A study of the rate of recrystallization when the specimen had not received such heating will now be described.

The preliminary crystallization was first carried out overnight at -33°C according to the standard procedure already described. The specimen was then plunged into a bath held at 0°C and observations made of the change of volume with time. It will be noted from Fig. 4 that the melting of crystals formed at -33°C is complete below 0°C . At first there was a volume increase due to thermal expansion and to the melting of the crystals. This change was found to be complete within about 5 minutes, a reasonable time for the establishment of thermal equilibrium. After an interval of time there was a volume decrease due to recrystallization at 0° . The whole procedure was repeated with the preliminary

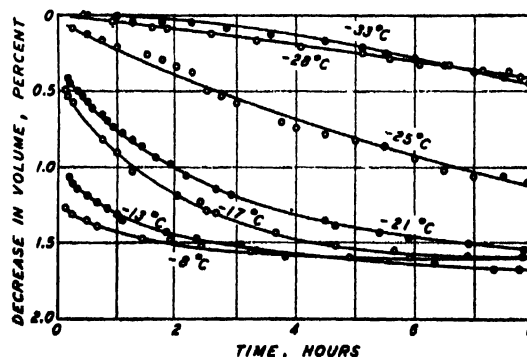


FIG. 5. Recrystallization of rubber at 0°C . The curves show the change of volume as a function of time, after preliminary crystallizations at the fixed temperatures indicated on the curves. The zero on the ordinate axis represents the volume of amorphous rubber at 0°C . The zero on the time scale represents the time at which the dilatometer was plunged into the bath at 0°C . Observations taken in the first five minutes, before the attainment of temperature equilibrium, are omitted. The upper ends of the melting ranges as read from Fig. 4, corresponding to the different temperatures of preliminary crystallization are -2° , -1° , $+1^{\circ}$, $+2^{\circ}$, $+4^{\circ}$, $+6^{\circ}$, and $+9^{\circ}\text{C}$, respectively.

crystallization occurring at -28° , -25° , -21° , -17° , -13° , and -8°C , respectively. The results are reproduced in Fig. 5.

Very similar results were obtained when the conditions were altered so that the preliminary crystallization was carried out in each case at -18°C and the recrystallization at temperatures of -8° , -4° , 0° , $+1^{\circ}$, and $+4^{\circ}\text{C}$, respectively. The results are shown in Fig. 6. The significance of these results is discussed in a later section.

5. Stark Rubber

When received in temperate climates, crude rubber is sometimes found to be crystallized^{24,29} and to have a melting range above room temperature. For convenience in the present discussion, crystalline rubber having a melting range extending above 25°C will be called stark rubber in accordance with the suggestion of Pickles⁶¹ and Whitby⁷² since it is stiff and rather rigid at room temperature. The use of such a term is not intended to imply any essential fundamental difference between stark rubber and that melting at lower temperatures. It is to be presumed on the basis of Fig. 4 that during the formation of the crystals in stark rubber conditions were such that the stark rubber was crystallized at temperatures higher than about 10°C . No cases, other than those in the present paper, seem to have been reported in which the temperature of crystallization was measured (or even estimated, within 10°C) during the formation of stark rubber.

One sample of stark rubber, for which the authors are indebted to Professor H. I. Cramer, formerly of the University of Akron, was found to have a melting range from about 32° to 39°C . When specimens of this stark rubber were placed in a dilatometer and held at a temperature of about 1°C for 48 days considerable additional crystallization occurred, most of it within the first few days. When the specimens were melted in the usual manner the curves shown in Fig. 7 were obtained. The line between 16° and 32°C passes through the points obtained before the stark rubber was held at 1°C for the additional crystallization.

In another experiment one end of a strip of stark rubber was immersed in water at 50°C for a few minutes. The lower end was thereby melted

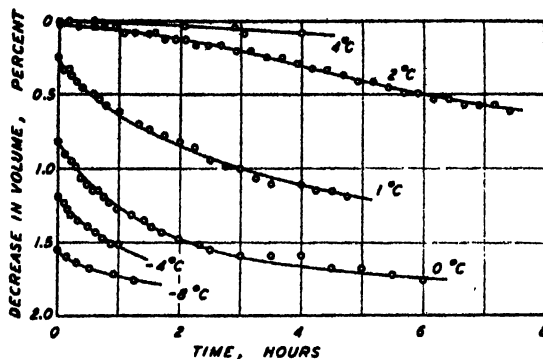


FIG. 6. Recrystallization of rubber after preliminary crystallization at -18°C . The curves show the change of volume as a function of time at the temperatures indicated. The zero on the ordinate axis represents, in each case, the volume of amorphous rubber at these temperatures. The zero on the time scale represents the time at which the dilatometer was plunged into the bath at the temperatures indicated. The dark circles represent the volume observed immediately after the attainment of thermal equilibrium. The upper end of the melting range, as read from Fig. 4, corresponding to preliminary crystallization at -18°C is $+4^{\circ}\text{C}$.

while the upper end remained in the stark condition. When the whole strip was placed in a refrigerator at 1°C for several weeks, the lower end crystallized and could not be distinguished by visual or tactual observation from the upper end. However, when the strip was allowed to come to room temperature, the crystals in the lower end melted, as would be expected from Fig. 4, while the upper end remained stark.

IV. DISCUSSION OF RESULTS

1. The Rate of Crystallization

It should be recognized that the measurements reported here regarding the rate of crystallization were all made on not more than two or three specimens of smoked sheet which had been heated to 55°C immediately before the crystallization. The conditions of preparation of the rubber from the latex undoubtedly have a considerable influence on the rate of crystallization.⁴² Uncoagulated latex itself showed no volume change when kept in a dilatometer at 2°C over a period of several months. This behavior confirmed previous observations in this laboratory.⁵ Likewise, qualitative observation showed that crystallization was extremely slow in a sheet of evaporated latex. In this case crystallization appeared to take place most readily in certain regions. This effect is similar to the "patchy

crystallization" described by Treloar,⁶⁹ who reported that visual observations on a sample of smoked sheet showed that neighboring portions crystallized at widely different rates. No attempt was made in the present work to investigate the exact reasons for such differences. Various factors which have been reported to increase the rate of crystallization are residual stresses remaining after previous mechanical treatment, moderate pressure^{40, 65, 66, 77} and exposure to a beam of x-rays.⁴⁷ Very high pressures may inhibit crystallization.¹⁸ The effect of previous thermal treatment on the rate of crystallization is discussed in the sections dealing with recrystallization.

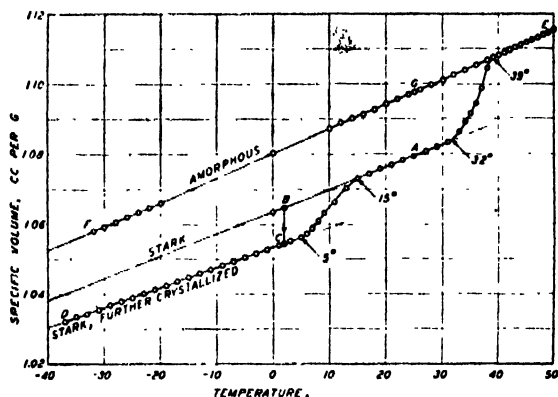


FIG. 7. Further crystallization and melting of a sample of stark rubber. The specific volume of the sample as received is shown at A. It was cooled to B, where further crystallization occurred, reducing the specific volume to C. After cooling to D the temperature was raised to E. The temperature was lowered to F and then raised to room temperature at G.

It would appear reasonable from theoretical considerations to suppose that an increase of mobility of the atomic chains, brought about for example by mastication of the rubber, would increase the rate of crystallization. Although van Rossem and Lotichius⁵³ could find only insignificant differences in the rate of crystallization with increasing times of milling, Katz,⁵⁴ Gehman,⁵⁵ and Cotton¹⁷ all state that masticated rubber crystallizes more readily than unmasticated. The reverse effect, a decrease in the rate of crystallization caused by a decrease in the mobility of the chains on vulcanization has already been described.⁸

The rates of crystallization observed in the present work are in reasonable agreement with observations of density by Bekkedahl¹⁵ and Treloar⁶⁹ at 0°C, by Smith and Hanna⁵⁶ at

-20°C, and by van Rossem and Lotichius⁵³ at 4°C and -10°C. Likewise a similar rate can be calculated from observations of Young's modulus at 0°C made by Conant and Liska.¹⁶

2. Lower Limit of Temperature of Crystallization

It has been recognized for some years that rubber does not crystallize at extremely low temperatures. No crystallization was evident in a specimen⁵ cooled at -259°C, or in another⁵² held at -190°C for 8 days. Other specimens^{5, 49} were kept between -50°C and -78°C for 3 weeks without the occurrence of crystallization but did crystallize at -35°C and above. The present work serves to locate more definitely the lower limit of temperature at which crystallization is possible as about -50°C. Figure 2 shows that the rate of crystallization has become negligible below this temperature. Here the mobility is presumably insufficient to permit the orientation necessary for the formation of crystals. The supercooling of rubber is similar in this respect to that of selenium,⁴⁶ sulfur, and many organic liquids.

3. Upper Limit of Temperature of Crystallization

Crystallization has not been reported at the room temperatures (usually above 20°C) normally maintained in American laboratories. No evidence of it was noted in the present investigation when dilatometers were kept at room temperature over a period of a year or more. In European laboratories, however, where the temperatures are frequently near 15°C, the crystallization of rubber on storage has been noted on a number of occasions.^{38, 44, 46, 52} Cotton,¹⁷ for example, states: "When smoked sheet and crepe rubber have been in storage for some time, they gradually harden and assume a frozen condition. This is due to crystallization." In the present work the highest temperature at which crystallization was observed was 14°C. At this temperature under the experimental conditions described here the time required for crystallization is of the order of years. Crystallization at this temperature was continuing at an approximately constant rate when the experiment was discontinued after 289 days.

The crystallization observed in European laboratories appears to occur at a more rapid rate than this. Quite probably temperatures lower than 15°C during the night initiated the crystallization and the crystals then formed were not melted near 15°C during the day.

It is possible to increase the rate of crystallization by varying the previous thermal or mechanical treatment of the specimen, and thereby forming more centers of crystallization. Consequently, it is likely that the upper limit of temperature of crystallization shown in Fig. 2 may be raised somewhat by the proper previous treatment of the rubber.

4. Temperature at which Crystallization is Most Rapid

The rate of crystallization was found to be a maximum at a temperature near -25°C, as can be seen in Fig. 2. In the production of the form of raw rubber known as "cut sheet" or "patent rubber,"²⁷ developed in Europe many years ago, rubber is crystallized so that it can be cut into sheets. The rubber is often held at a temperature of about -5° for a few days in order to bring about the crystallization. More recent books^{3, 64} state that the rubber should be held between -5° and -10°C for at least six days. No quantitative study of the rates at different temperatures, as in the present work, seems to have been reported previously.

5. Magnitude of the Volume Change on Crystallization

The magnitude of the percentage decrease of volume on crystallization can be noted from curves of the type shown in Fig. 1. The actual specific volume of the amorphous rubber varies with temperature, and can be read from curves like those in Fig. 3. The decrease of specific volume on crystallization is of course found to be the same as the increase of specific volume on subsequent melting. However it might be pointed out that the percentage decrease of volume is not quite the same as the subsequent percentage increase of volume, since they are fractions of different quantities. Furthermore, any comparisons must be made at the same temperature.

In the present work in most cases the decreases

of volume on crystallization were found to lie between 2.0 and 2.7 percent. Repeated crystallizations of the same specimen under conditions which were thought to be identical did not always lead to the same volume changes. It seems possible that such variations are statistical fluctuations occasioned by the growth of crystals from different centers in different cases. There was an even greater variation in the decreases of volume when different specimens were studied but in almost all cases the values lay within the limits given. The exact value, within these limits, seems to depend upon factors which were not under control.

Values for this quantity, calculated from measurements of density or specific gravity reported by previous workers^{4, 5, 8, 14, 19, 31, 39, 52, 55, 56, 72} generally lie within the same limits, although a few are somewhat lower. Holt and McPherson²⁸ found a volume change of about 1.85 percent on stretching a specimen of vulcanized rubber to an elongation of 700 percent at 25°C. Treloar⁶⁹ by stretching unvulcanized rubber at 0°C, observed a volume decrease of over 3 percent, the highest value reported. With unstretched pale crepe he found a volume decrease of 2.3 percent on crystallization at 0°C.

Observations made in the course of the present work furnish no data from which conclusions may be drawn as to the relative amounts of crystalline and amorphous material in a specimen.

Field^{21, 24} has recently conducted x-ray studies of this proportion of crystalline and amorphous components in stretched rubber, and concludes that under favorable conditions about 80 percent of the material is in the crystalline state, in agreement with an earlier study by Meyer and Mark.⁴³ This value is considerably higher than those estimated by Parks⁵⁰ and Wildschut⁷³ by other methods. If reliance is to be placed on x-ray values, it would be very desirable to conduct several parallel studies of crystallization by means of x-rays and by means of observations of volume change. In this manner the relation between the percentage decrease of volume and percentage of crystalline material could be established.

6. Melting of Crystallized Rubber

The melting of the crystalline rubber is very much dependent on the temperature at which the

TABLE I. Melting temperatures of crystalline rubber previously reported.

Lower limit °C	Melting Maximum change °C	Upper limit °C	Crystallization temperature °C	Observers	Technique	Remarks
	43.5			Whitby ^{7a}	Density	
	43			Whitby ^{7a}	Density	
	41			Whitby ^{7a}	Density	
	41			Barnes ⁴	Visual	{Flexibility Transparency
40				Wood, Bekkedahl, and Peters ^{7b}	Linear expansion	
40	up to 37	50	5 to 15	Feuchter ¹⁹	Density	After 1 year
	36			Katz ²⁶	X-ray	
	37			Katz ²⁷	X-ray	"old rubber"
				van Rossem and Lotichius ²⁸	Density, hardness, and light absorption	
35		45		Meyer, von Susich, and Valko ⁴⁴	X-ray	
35				Wood, Bekkedahl, and Peters ^{7b}	Linear expansion	
34	36	>40		van Rossem and Lotichius ²⁸	Density	Stored 13 years
	35			Feuchter and Hauser ²⁰	Retraction	1 year after racking
		35 to 40		von Susich ⁴⁰	X-ray	
31.5		33		Cotton ¹⁷		"Normally frozen"
30		40		Feuchter ¹⁹	Density	Masticated
30		39		Katz ²⁶	Density	
30	33	35		Pickles (van Rossem) ⁵¹	Density	
28	32	35		van Rossem and Lotichius ²⁸	Density	
27		29		Katz ²⁶	X-ray	"Cut sheet"
25		36		Feuchter and Hauser ²⁰	Retraction	Racked
20		30		Katz ²⁶	X-ray	
20		30		Bunschoten ¹⁴	Density	

crystals have been formed, as can be seen from Figs. 3 and 4. This conclusion seemed so surprising that it was made the subject of a preliminary communication⁷ based upon only a part of the data reported in the present paper.

It can be seen that the higher the temperature of crystallization the higher are the temperatures at which both the beginning and ending of the melting occur. The beginning of melting occurs at a temperature of from 4° to 7° above that at which the crystals have been formed. For all temperatures of crystallization below about -35°C the temperature at which melting is complete is about -2°C.

The range of melting becomes narrower the higher the temperature of crystallization up to about 0°C; for crystals formed at temperatures between 0° and 14°C, the range of melting is about 10°; for crystallization at higher temperatures no direct observations are available, but the

range is probably narrower, since it appears that stark rubber has a somewhat narrower range of melting the higher the mean temperature of the range.

The results presented here undoubtedly explain many previously reported discrepancies in the temperatures of melting of crystalline rubber. Unfortunately in only a few cases have previous workers specified the temperature at which crystallization occurred. The variations in the melting have sometimes been thought to be connected with the variety of rubber, or the experimental technique. In most cases no explanation at all was attempted. Table I presents the results of measurements of the melting temperatures of crystalline rubber, as reported by previous observers. In those instances in which the temperature of crystallization was reported the only pronounced deviations from the results of Fig. 4 are in the few instances where the crystals were

TABLE 1.—*Continued.*

Lower limit °C	Melting		Crystallization temperature °C	Observers	Technique	Remarks
	Maximum change °C	Upper limit °C				
20	25	30	Cool cellar	Ruhemann and Simon ⁴⁴	Specific heat	Smoked sheet
		<30	—5.8	Carson ¹⁸		
20		24		Hock ²¹	Retraction	
20		23		Hock ²⁰	Retraction	
20		23	7 to 10	Thiessen and Kirsch ⁴⁶	X-ray	Crystallized under pressure of 10–25 atmospheres
10	18	30	—15	Ruhemann and Simon ⁴⁴	Specific heat	"Cut sheets" crystallized at 150 atmospheres
11		13	6	Thiessen and Kirsch ⁴⁶	X-ray	Pressure of 30 atmospheres
9.5		11	—58 to —43	Smith, Saylor and Wing ⁴⁹	Birefringence	Sol rubber crystallized from solution
8		10		Cotton ¹⁷	Visual	Frozen in refrigerator
		<20	0 to 2	Katz and Bing ³⁸	X-ray	
		<20	—5 to —10	Katz ³⁴	X-ray	
		<20	—10	Katz and Bing ³⁸	X-ray	
6	11	17	2	Bekkedahl and Wood ⁸	Volume	Vulcanized rubber
6	11	16	0	Bekkedahl ⁴	Volume	
	10			Park ⁴⁸	—	
2		14	—10	van Rossem and Lotichius ⁴⁵	Density and hardness	
0	5	10		Ruhemann and Simon ⁴⁴	Specific heat	
—2		14	—50	Smith and Saylor ⁴⁷	Birefringence	Gel crystallized from solution
—5.5		16	—25	Smith and Saylor ⁴⁸	Birefringence	
—15	11	17	<0	Bekkedahl and Matheson ⁶	Specific heat	
	0			Bunn and Garner ¹³	—	
—20		10		Ruhemann and Simon ⁴⁴	Specific heat	Beta-anomaly

formed from solution rather than from the bulk material.

A study of the table confirms the conclusion of the present investigation that there are not merely two forms of crystalline rubber, as has sometimes been suggested^{4,17} but rather that crystals melting at any temperature between about —40° and +45°C may be formed in rubber by the proper choice of temperature for crystallization. It is rather surprising that such a conclusion has not been already drawn on the basis of the experimental data shown in Table I. The assignment of a definite melting point or melting range to rubber in general is thus seen to be without meaning in spite of the many investigators who have hoped to do so.^{5,13,17,30,44,48,66} A recent writer¹² on the subject expresses considerable wonder at the variability of the melting point.

The temperature at which retraction of stretched rubber occurs is sometimes associated with the melting of crystals formed on stretching. Although the scope of the present investigation did not include any measurements on stretched rubber some of the results may probably be safely applied to crystals formed by stretching. For example the influence of the temperature of crystallization on the melting seems to furnish an explanation of certain effects observed by Treloar⁶⁶ in the retraction of stretched rubber. He found that the recovery temperature was dependent on the temperature at which stretching was performed. Lacking the results reported here, he said "One would not expect the melting point of the crystals to depend on the temperature at which they were formed."

Some years ago van Rossem and Lotichius⁴⁵

made a rather careful study of the crystallization of raw rubber. In the main, there seems to be no disagreement between their experimental results and those contained in the present paper. In one instance it now seems desirable to give to their results an interpretation different from the one originally given.

This instance, which is one of the best-known of the conclusions of van Rossem and Lotichius, is a statement that the melting temperature of one sample of stark rubber was 31° – 33°C in 1919 and had become 35° – 37°C in 1927. An examination of the graph giving the specific gravity values which are the basis for discussion shows that the terms "melting temperature" or "melting point" were limited to the steepest portions of the curves. The full melting range of the specimen observed in 1919 appears to extend from 28° to 35°C . The observations made in 1927 show that the melting range began at 34° and extended above 40°C , the highest temperature at which observations were made. Since the specimen measured in 1927 was not completely melted at 40° , its specific gravity did not come down to the value characteristic of the amorphous rubber, as did that of the specimen measured in 1919, where melting was apparently complete at 35°C . No indications were given regarding the temperature at which the specimen was stored from 1919 to 1927. The observed change of melting range could be explained on the reasonable assumption that the temperature, at least part of the time, was between 28° and 34°C . Crystals melting below 34° would then be melted, and re-crystallization at those temperatures would yield new crystals having a higher melting range, as already described in the section on recrystallization.

Thus there seems to be no assurance that van Rossem and Lotichius would have found a rise in the melting range during storage at a constant temperature below 28°C . No such rise was observed in the course of the present work. The actual phenomena observed by van Rossem and Lotichius can then be explained in terms of the results of the present paper, without recourse to speculative assumptions of changes in crystal size or polymerization.

The considerations developed in the present work on the melting of crystalline rubber have already been applied⁷ to explain results obtained

by Ruhemann and Simon⁸ in measurements of specific heat. It has been pointed out that the beta-anomaly which they observed can be explained as a phenomenon of crystallization under the particular conditions of their experiments, without the assumption of an anomaly characteristic of rubber in general. The volume-temperature graphs which they give for rising temperatures are of the same form as those obtained in the present work when the rate of temperature rise was slow enough to permit recrystallization and subsequent melting at higher temperatures.

7. Recrystallization

A number of interesting conclusions may be drawn from the experiments on recrystallization illustrated in Figs. 5 and 6. In the first place the rate of crystallization at a given temperature is much greater than in the instances previously discussed, where the standard procedure called for heating to at least 55°C before each crystallization to destroy possible centers of crystallization. An examination of Fig. 1 shows that at 2°C after such treatment the change of volume in the first 10 hours is about 0.05 percent. On the scales used in Figs. 5 and 6 such a change would hardly be perceptible and the experimental points would lie on an almost horizontal line through the origin. In the second place the effects of the preliminary crystallization in increasing the rate of subsequent crystallization extend at least several degrees above the upper end of the melting range. For example it is clear from Fig. 6 that the melting range does not extend above 4°C , since the initial ordinate is zero for the recrystallization at 4°C . The rate of recrystallization at 4°C is seen to be of appreciable magnitude in comparison with the much lower rate which is observed when the centers of crystallization are destroyed by previous heating to 55°C .

The selection of a temperature of 55°C as being sufficient to destroy centers of crystallization was based on the fact that it was five or ten degrees above the highest temperature at which any crystals have been reported in unstretched rubber, as noted in Table I. In numerous experiments no evidence of any effect of previous thermal history was found persisting in any sample after heating it to 55°C .

Attention should be called to the two different types of crystallization curves shown in Figs. 5 and 6. At temperatures below the upper end of the melting range the rate of crystallization has its largest value at the beginning and decreases with increasing time. At temperatures above the upper end of the melting range the rate of crystallization is small at the beginning, increases with time to a maximum, and, when observations are carried out over a sufficiently long period, decreases again to a negligible value. The latter curves are thus of the sigmoid type shown in Fig. 1. One can therefore conclude from Fig. 5 alone that the end of the melting range would be at 0°C following crystallization at a temperature between -25° and -28°C. From Fig. 4 this temperature is read as -28°C.

In terms of the usual conceptions of crystal growth it is possible to give a simple explanation of the difference between the two types of curves. In accordance with the ideas of Tammann and others^{3, 23, 24, 33, 61, 62, 71} a crystal can be initiated only at certain centers or nuclei. The crystal, once initiated, grows from the center at a characteristic rate, called the linear crystallization velocity, until the available crystallizable material between it and the next crystal is exhausted.

In the first instance where crystallization represents merely further growth of crystals already present, no additional centers of crystallization are necessary. The rate of crystallization as measured by volume change is therefore a maximum at the beginning and decreases because of the exhaustion of crystallizable material. In the second instance crystallization at the beginning is slow because of the lack of a sufficient number of centers from which crystals may grow. Crystals start from the available centers, and more centers are formed as time goes on.^{3, 24, 76} Thus the rate of crystallization increases at first, reaching a maximum, and then decreases again with the exhaustion of crystallizable material.

At temperatures immediately above the upper end of the melting range there remains of course no crystalline material. However, centers of crystallization must persist at least several degrees above the end of the melting range, since the rate of recrystallization at these temperatures, as shown in Fig. 7, is much greater than is

the case when the centers of crystallization are destroyed by heating to 55°C.

8. Nature of the Crystals

The nature of the crystalline regions in long chain high polymers has been the subject of considerable discussion in recent years. Alfrey and Mark,^{1, 2, 41} Bunn,¹⁰⁻¹² Treloar,⁷⁰ and Wood⁷⁸ have summarized some of the ideas. It seems generally agreed, for instance, that the structural units of the crystallites are not whole molecules but rather kinetic units which are segments of chains. Consequently, a single chain may contain segments which are units in two or more different crystallites, joined by segments which are to be regarded as part of the amorphous material.

The scope of the present paper does not include any attempt to extend the detailed conception of the nature of the crystals on the basis of the results obtained here. It has already been pointed out elsewhere⁷⁸ that these results cast grave doubts on the acceptability of several theories proposed to account for the existence of a range of melting. The field is still open for a theory which will logically and satisfactorily explain two experimental observations reported here: (1) that the range of melting is not dependent on the extent of crystallization and (2) that it is definitely determined by the temperature at which crystallization occurs.

Crystals of a lower-melting type can be formed and melted in a specimen already containing crystals of a higher-melting type, as shown in the experiments illustrated by Fig. 7. It is rather remarkable that the presence of crystals of the higher-melting type appears to have so little effect on the formation and melting of the lower-melting crystals. It does appear that the two types must compete for crystallizable material, since the total volume change does not exceed 3 percent in any case.

It cannot be emphasized too strongly that the sole important factor in determining the melting range is the temperature at which crystallization occurs. Other conditions may alter the rate of crystallization by a large factor or the volume change during crystallization, as already discussed, but the melting range remains the same. Even different degrees of vulcanization⁸ do not

alter the relation between melting range and temperature of crystallization. Thus, the quantitative results depicted in Fig. 4 are regarded to be of much more general application than the other portions of the present investigation.

The experimental work described in this paper was performed between 1940 and 1942, but war activities prevented earlier publication, except for a preliminary note⁷ on the change of melting range with temperature of crystallization.

BIBLIOGRAPHY

- (1) T. Alfrey and H. Mark, "Phase transitions in rubber and their connection with the stress-strain curve," *Rubber Chem. Tech.* **14**, 525 (1941).
- (2) T. Alfrey and H. Mark, "A statistical treatment of crystallization phenomena in high polymers," *J. Phys. Chem.* **46**, 112 (1942); Reprinted in *Rubber Chem. Tech.* **15**, 462 (1942).
- (3) M. Avrami, "Kinetics of phase change," *J. Chem. Phys.* **7**, 1103 (1939); **8**, 212 (1940); **9**, 177 (1941).
- (4) W. H. Barnes, "Diffraction of x-rays by old specimens of 'frozen' rubber," *Can. J. Research* **15**, 156 (1937); Reprinted in *Rubber Chem. Tech.* **10**, 462 (1937).
- (5) N. Bekkedahl, "Forms of rubber as indicated by temperature-volume relationship," *J. Research Nat. Bur. Stand.* **13**, 411 (1934). RP717; Reprinted in *Rubber Chem. Tech.* **8**, 5 (1935).
- (6) N. Bekkedahl and H. Matheson, "Heat capacity, entropy, and free energy of rubber hydrocarbon," *J. Research Nat. Bur. Stand.* **15**, 503 (1935), RP844; Reprinted in *Rubber Chem. Tech.* **9**, 264 (1936).
- (7) N. Bekkedahl and L. A. Wood, "Influence of the temperature of crystallization on the melting of crystalline rubber," *J. Chem. Phys.* **9**, 193 (1941); Reprinted in *Rubber Chem. Tech.* **14**, 544 (1941).
- (8) N. Bekkedahl and L. A. Wood, "Crystallization of vulcanized rubber," *Ind. Eng. Chem.* **33**, 381 (1941); Reprinted in *Rubber Chem. Tech.* **14**, 347 (1941).
- (9) S. Boström, K. Lange, H. Schmidt, and P. Stöcklin, *Kautschuk und verwandte Stoffe* (Union Deutsche Verlagsgesellschaft Berlin Roth and Company, Berlin, 1940).
- (10) C. W. Bunn, "Molecular structure and rubber-like elasticity. I. The crystal structure of *B* gutta-percha, rubber, and polychloroprene," *Proc. Roy Soc. London* **A180**, 40 (1942); Reprinted in *Rubber Chem. Tech.* **15**, 709 (1942).
- (11) C. W. Bunn, "Molecular structure and rubber-like elasticity. III. Molecular movements in rubber-like polymers," *Proc. Roy Soc. London* **A180**, 82 (1942); Reprinted in *Rubber Chem. Tech.* **15**, 742 (1942).
- (12) C. W. Bunn, "The stereochemistry of the rubber molecule," *Trans. Faraday Soc.* **38**, 372 (1942); Reprinted in *Rubber Chem. Tech.* **15**, 704 (1942); **16**, 775 (1943).
- (13) C. W. Bunn and E. V. Garner, "The crystal structure of rubber hydrochloride," *J. Chem. Soc.* **1942**, 654 (1942); Reprinted in *Rubber Chem. Tech.* **16**, 848 (1943).
- (14) B. Bunschoten, Mededeel Rijksvoorlichtingsdienst Behoeve Rubberhandel Rubbernijverheid, Delft, No. 3, 37 (1921).
- (15) C. M. Carson, "Effect of storage on milled crude rubber," *Ind. Eng. Chem.* **23**, 691 (1931); Reprinted in *Rubber Chem. Tech.* **4**, 417 (1931).
- (16) F. S. Conant and J. W. Liska, "Some low-temperature properties of elastomers," *J. App. Phys.* **15**, 767 (1944); Reprinted in *Rubber Chem. Tech.* **18**, 318 (1945).
- (17) F. H. Cotton, "Rational processing," *Trans. Inst. Rubber Ind.* **16**, 303 (1941); Reprinted in *Rubber Chem. Tech.* **14**, 762 (1941).
- (18) R. B. Dow, "Inhibition of crystallization of rubber by high pressure," *J. Chem. Phys.* **7**, 201 (1939); Reprinted in *Rubber Chem. Tech.* **12**, 496 (1939).
- (19) H. Feuchter, "Concerning racked rubber, its 'melting line,' and its density," *Kautschuk* **2**, 260 and 282 (1926).
- (20) H. Feuchter and E. A. Hauser, "Concerning aging, hysteresis, and melting point as well as work capacity and work power of progressively racked rubber," *Kautschuk* **5**, 194, 218, 245, and 276 (1929).
- (21) J. E. Field, "An x-ray study of the proportion of crystalline and amorphous components in stretched rubber," *J. App. Phys.* **12**, 23 (1941); Reprinted in *Rubber Chem. Tech.* **14**, 555 (1941).
- (22) S. D. Gehman, "The contribution of x-ray research to the knowledge of rubber," *Chem. Rev.* **26**, 203 (1940).
- (23) S. D. Gehman and J. E. Field, "An x-ray investigation of crystallinity in rubber," *J. App. Phys.* **10**, 564 (1939); Reprinted in *Rubber Chem. Tech.* **12**, 706 (1939).
- (24) S. D. Gehman and J. E. Field, "Observations on the x-ray structure of rubber and the size and shape of rubber crystallites," *J. App. Phys.* **15**, 371 (1944); Reprinted in *Rubber Chem. Tech.* **17**, 640 (1944).
- (25) A. Goetz and S. S. Goetz, "Vitrification and crystallization of organic cells at low temperatures," *J. App. Phys.* **9**, 718 (1938).
- (26) F. von Goler and G. Sachs, "The kinetics of crystallization phenomena," *Zeits. f. Physik* **77**, 281 (1932).
- (27) K. Gottlob, *Technology of Rubber*, Translated into English by J. L. Rosenbaum (MacLaren Sons. Ltd., London, 1927), p. 319.
- (28) E. A. Hauser, M. Hunemörder, and P. Rosbaud, "New x-ray studies on rubber and related substances," *Kautschuk* **3**, 228 (1927).
- (29) E. A. Hauser and P. Rosbaud, "X-ray studies on rubber and similar substances," *Kautschuk* **3**, 17 (1927).
- (30) L. Hock, Chapter on "Physics of Rubber" in *Memmler's Science of Rubber*, Translated by R. F. Dunbrook and V. N. Morris (Reinhold Publishing Corporation New York, New York, 1934), p. 421.
- (31) L. Hock, "Theory of the Joule effect in rubber," *Zeits. f. Elektrochemie* **31**, 404 (1925).
- (32) W. L. Holt and A. T. McPherson, "Change of volume of rubber on stretching; effects of time, elongation, and temperature," *J. Research Nat. Bur. Stand.* **17**, 657 (1936), RP936; Reprinted in *Rubber Chem. Tech.* **10**, 412 (1937).
- (33) W. A. Johnson and R. F. Mehl, "Reaction kinetics in processes of nucleation and growth," *Trans. Am. Inst. Min. Met. Eng.* **135**, 416 (1939).
- (34) J. R. Katz, "X-ray spectrographic investigations of stretched rubber and their possible significance in the problem of the elastic properties of this material," *Naturwiss.* **13**, 410 (1925).
- (35) J. R. Katz, "Crystal interferences in rubber and other polyrenes," *Gummi-Ztg.* **41**, 1 (1927).
- (36) J. R. Katz, "The freezing of raw rubber. X-ray part," *Kautschuk* **5**, 6 (1929).
- (37) J. R. Katz, "X-ray spectrography of polymers and in particular those having a rubber-like extensibility," *Trans. Faraday Soc.* **32**, 77 (1936); Reprinted in *Rubber Chem. Tech.* **9**, 357 (1936).
- (38) J. R. Katz and K. Bing, "Is raw rubber partially crystalline?" *Zeits. f. angew. Chemie* **38**, 439 (1925).

- (39) F. Kirchhof, "Raw rubber—a fibrous material," *Kautschuk* 2, 151 (1926).
- (40) W. Kirsch, "Equilibria between crystals and melts in rubber and their dependency on temperature and pressure," Dissertation (Trillsch and Huther, Berlin, 1937).
- (41) H. Mark, "Phase transitions and elastic behavior of high polymers," *Ind. Eng. Chem.* 34, 449 (1942); Reprinted in *Rubber Chem. Tech.* 15, 452 (1942).
- (42) T. H. Messenger and J. R. Scott, "Softening and freezing of crepe," *J. Research Assn. Brit. Rubber Mfrs.* 5, 64 (1936); Reprinted in *Rubber Chem. Tech.* 10, 7 (1937).
- (43) K. H. Meyer and H. Mark, "Rubber," *Ber. d. d. chem. Ges.* 61, 1939 (1928).
- (44) K. H. Meyer, G. von Susich, and E. Valko, "The elastic properties of the organic high polymers and their kinetic significance," *Kolloid Zeits.* 59, 208 (1932).
- (45) P. Mondain-Monval, "Investigations of the allotropy and the crystallization of vitreous materials," *Ann. de chim.* 3, 5 (1935).
- (46) E. Ott, "Concerning the molecular size of rubber and gutta percha," *Naturwiss.* 14, 320 (1926).
- (47) E. Ott, "X-ray investigations on rubber," *J. Am. Chem. Soc.* 52, 4612 (1930); Reprinted in *Rubber Chem. Tech.* 4, 82 (1931).
- (48) C. Park, "On the crystallization of crude rubber," *Memoirs of College of Science, Kyoto Imperial University* A22, 13 (1939). *J. Soc. Rubber Ind. Japan* 12, 462 (1939) (in English); Reprinted in *Rubber Chem. Tech.* 12, 778 (1939).
- (49) C. Park, "Second report of the investigation of the crystallization of crude rubber," *Memoirs of the College of Science Kyoto Imperial University* A22, 259 (1939) (in English); Reprinted in *Rubber Chem. Tech.* 13, 539 (1940).
- (50) G. S. Parks, "The amorphous and crystalline forms of rubber hydrocarbon," *J. Chem. Phys.* 4, 459 (1936); Reprinted in *Rubber Chem. Tech.* 10, 135 (1937).
- (51) S. S. Pickles, "The consistency of rubber and rubber compounds," *India Rubber J.* 67, 69 (1924).
- (52) R. Pummerer and G. von Susich, "Crystallized rubber," *Kautschuk* 7, 117 (1931); Translated in *Rubber Chem. Tech.* 5, 245 (1932).
- (53) A. van Rossem and J. Lotichius, "The 'freezing' of raw rubber," *Kautschuk* 5, 2 (1929); Translated in *Rubber Chem. Tech.* 2, 378 (1929).
- (54) M. Ruhemann and F. Simon, "A contribution to the knowledge of the physical properties of rubber," *Zeits. f. physik. Chemie* A138, 1 (1928).
- (55) H. Sachse, "The thermal effects accompanying the spontaneous density increase in raw rubber," *Gummi-Ztg.* 43, 2216 (1929).
- (56) W. H. Smith and N. P. Hanna, "Comparison between the observed density of crystalline rubber and the density calculated from x-ray data," *J. Research Nat. Bur. Stand.* 27, 229 (1941), RP1416; Reprinted in *Rubber Chem. Tech.* 15, 265 (1942).
- (57) W. H. Smith and C. P. Saylor, "Ether-insoluble or gel rubber hydrocarbon, its solution, crystallization, and properties," *J. Research Nat. Bur. Stand.* 13, 453 (1934), RP719; Reprinted in *Rubber Chem. Tech.* 8, 214 (1935).
- (58) W. H. Smith and C. P. Saylor, "Optical and dimensional changes which accompany the freezing and melting of Hevea rubber," *J. Research Nat. Bur. Stand.* 21, 257 (1938), RP1129; Reprinted in *Rubber Chem. Tech.* 12, 18 (1939).
- (59) W. H. Smith, C. P. Saylor, and H. J. Wing, "Preparation and crystallization of pure ether-soluble rubber hydrocarbon: composition, melting point, and optical properties," *Bur. Stand. J. Research* 10, 479 (1933), RP544; Reprinted in *Rubber Chem. Tech.* 6, 351 (1933).
- (60) G. von Susich, "The fusion curve of natural rubber," *Naturwiss.* 18, 915 (1930); Translated in *Rubber Chem. Tech.* 4, 211 (1931).
- (61) G. Tammann, *Der Glaszustand* (Leopold Voss, Leipzig, 1933).
- (62) G. Tammann and E. Jenckel, "The crystallization velocity and the number of nuclei in glycerine and their dependence on temperature," *Zeits. f. anorg. allgem. Chemie* 193, 76 (1930).
- (63) W. E. Thibodeau and A. T. McPherson, "Photoelastic properties of soft vulcanized rubber," *J. Research Nat. Bur. Stand.* 13, 887 (1934), RP751; Reprinted in *Rubber Chem. Tech.* 8, 183 (1935).
- (64) F. Thiele and E. Wurm, "Chapter on Patentgummi" in E. A. Hauser's *Handbuch der Gesamten Kautschuk-technologie* (Union Deutsche Verlagsgesellschaft, Berlin, 1935), Vol. II.
- (65) P. A. Thiessen and W. Kirsch, "Crystallization of rubber by pressure," *Naturwiss.* 26, 387 (1938); Translated in *Rubber Chem. Tech.* 12, 12 (1939).
- (66) P. A. Thiessen and W. Kirsch, "The crystallization of weakly-vulcanized rubber by pressure," *Naturwiss.* 27, 390 (1939); Translated in *Rubber Chem. Tech.* 13, 48 (1940).
- (67) P. A. Thiessen and W. Wittstadt, "Forced and spontaneous changes in the arrangement of the molecules in stretched rubber. Crystals and fused phase in stretched rubber," *Zeits. f. physik. Chemie* B41, 33 (1938); Translated in *Rubber Chem. Tech.* 12, 736 (1939).
- (68) L. R. G. Treloar, "Elastic recovery and plastic flow in raw rubber," *Trans. Faraday Soc.* 36, 538 (1940); Reprinted in *Rubber Chem. Tech.* 13, 795 (1940).
- (69) L. R. G. Treloar, "Crystallization phenomena in raw rubber," *Trans. Faraday Soc.* 37, 84 (1941); Reprinted in *Rubber Chem. Tech.* 15, 251 (1942).
- (70) L. R. G. Treloar, "The structure and elasticity of rubber," *Reports on Progress in Physics* (Physical Society, London, 1943), Vol. 9, p. 113.
- (71) M. Volmer and M. Marder, "The theory of linear crystallization velocity of undercooled melts and undercooled solid modifications," *Zeits. f. physik. Chemie* A154, 97 (1931).
- (72) G. S. Whitby, Chapter 2 "Physical properties of raw rubber" in Davis and Blake's *Chemistry and Technology of Rubber* (Reinhold Publishing Corporation, New York, 1937), p. 102.
- (73) A. J. Wildschut, "On the proportion of crystalline and amorphous components in stretched vulcanized rubber," *J. App. Phys.* 17, 51 (1946).
- (74) W. Wittstadt, "External influence and the internal state of rubber," *Kautschuk* 15, 11 (1939); Translated in *Rubber Chem. Tech.* 12, 488 (1939).
- (75) L. A. Wood, "The beta-anomaly of Ruhemann and Simon in rubber," *J. Chem. Phys.* 10, 403 (1942).
- (76) L. A. Wood, "Crystallization Phenomena in Natural and Synthetic Rubbers," chapter in Mark and Whitby's *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. 2, p. 57.
- (77) L. A. Wood, N. Bekkedahl, and R. E. Gibson, "Effect of pressure on the melting of crystalline rubber," *J. Chem. Phys.* 13, 475 (1945); *J. Research Nat. Bur. Stand.* 35, 375 (1945), RP1677.
- (78) L. A. Wood, N. Bekkedahl, and C. G. Peters, "Application of the interferometer to the measurement of dimensional changes in rubber," *J. Research Nat. Bur. Stand.* 23, 571 (1939), RP1253; Reprinted in *Rubber Chem. Tech.* 13, 290 (1940).

Tackiness of GR-S and Other Elastomers

W. F. BUSSE, J. M. LAMBERT, AND R. B. VERDERY
General Aniline and Film Corporation, Easton, Pennsylvania

The deficiency of GR-S in "tackiness" resulted in a great deal of work in the rubber industry to improve this complex property, but practically all the work had to be evaluated by qualitative "hand tests" because quantitative tests were not available. An analysis of the factors involved in the "tackiness" of various materials showed that different factors were critical in the so-called "tackiness" of rubber, paints, and varnishes, printing inks, adhesive tapes, etc. A quantitative test was developed to measure the effects of rate of removal on the adhesion or "tackiness" of pressure-sensitive adhesive tapes over a 10⁴-fold range of rates. This showed why the results of standard "adhesion tests" of tapes do not predict the relative behavior of the

tapes under small forces acting for long times. Another test was developed for the measurement of the "tackiness" of GR-S rubber which correlates well with the judgment of tackiness by hand tests and it even works reasonably well with rough samples taken from a laboratory mill. With this test, the effects of mastication conditions, aging, and of the addition of softeners and tackifiers on "tackiness" of GR-S (with and without carbon black) were studied. The German tackifier "Koresin," which is now manufactured in this country, was found to be far more effective than rosin and similar materials formerly used, and gave a tackiness approaching that of natural rubber.

I. INTRODUCTION

TACKINESS usually is measured by hand tests which are very simple to apply but which are very complex because they involve an unconscious integration of a great many factors. The particular combination of properties which produces tack in one system, such as printing inks, may be quite different from the combination producing tack in another system such as rubber, even though the over-all effect as measured by hand tests may be the same. Hence it is not surprising that there is little agreement as to the factors involved in producing tackiness of different materials and that very few quantitative methods of measuring tackiness have been developed.

This paper presents an analysis of the operation of measuring tack, to make explicit some of the detailed factors involved, both in the materials and in the testing operation. Methods of measuring the effects of some of these different factors have been developed and applied to the study of the "tackiness" of such materials as pressure-sensitive adhesive tapes and various elastomers, including polyvinyl ethers and natural and synthetic rubbers.

The magnitude of the "tackiness" of a material (as measured by the resistance to the separation of two surfaces which have been brought in contact), depends on the geometry or roughness of the surface of the sample, as well as on the surface adhesive forces and the

visco-elastic properties of the material, and the relation of these factors to the time of application and magnitude of the force pressing the surfaces together. Hence, the results of quantitative tests of "tackiness" are not in any sense absolute properties of the material. However, they do give better measures of tackiness than can be obtained by hand tests, and they make possible more precise studies of the effects of some processing conditions and compounding agents on the tackiness of GR-S and natural rubbers.

II. HISTORY

Studies of the properties that produce "tack" in viscous liquids were made as early as 1847 by J. Stefan¹ and somewhat later by O. Reynolds.² Both considered the tackiness to be measured by the force required to separate, at a given rate, two flat disks immersed in the liquid, and they developed equations for relating the applied force to the times of separation, assuming the laminar flow of Newtonian liquids. Later similar formulas were developed,³ some of which⁴ were modified to take account of the yield value of the materials. Recently it has been shown⁵ that tack of printing inks correlates well with the viscosity measured at high rates of shear.

¹ J. Stefan, *Ber. Kais. Acad. Wiss. Wien* **69**, 713 (1847).
² O. Reynolds, *Phil. Trans. Roy. Soc. London* **177**, 157 (1886).

³ Haley, *Trans. Inst. Rubber Ind.* **1**, 334 (1926).

⁴ J. R. Scott, *Trans. Inst. Rubber Ind.* **17**, 169 (1931).

⁵ H. Green, *Ind. Eng. Chem. Anal. Ed.* **13**, 632 (1941).

Little work has been done relating tackiness to molecular properties, but it was pointed out⁶ that tackiness as well as "stickiness" and the ability of polymeric materials to be drawn into fibers, seems to be correlated with a relaxation time of 10^{-7} to 10^{-4} second, or a viscosity of 10^3 – 10^4 poise, and a linear chain length of about 100 links.

It has been emphasized⁷ that in the case of rubber "autohesion" or the sticking together of two like surfaces when brought in contact is an important factor in tackiness. The relations of various fundamental properties to the tackiness of paints, inks, rubbers, etc., were discussed by a number of authors at a recent meeting of the British Rheologists Club.⁸ In this discussion Scott emphasized two factors which are important in producing a bond between the two surfaces:⁹

- (1) getting the two surfaces into contact; and
- (2) mutual "fusion" of the surface layers.

Phase (1) of the process is greatly influenced by the roughness of the rubber surfaces. Phase (2) for rubber-like materials is thought to involve a rapid interpenetration, i.e., self-diffusion of at least parts of the long rubber molecules composing the two surface layers.

The rheology of the rubber is of great importance since it must deform readily (at least on the surface) under small stresses in order to produce close contact of the two surfaces. It is interesting that this requirement of ease of deformation for good tack is just opposite to the requirement for so-called tacky adhesion between the two plates separated by a layer of viscous or plastic material considered by Stefan and Reynolds.

Experience with GR-S shows that the addition of resinous materials and various other softeners to the rubber often makes it "sticky" to mill rolls, liners, etc., without giving it the combination of surface and bulk properties required for good autohesion or "tack." Extensive studies

have been made and published in this country¹⁰ on the effect of various softeners and "extenders" on the physical properties of GR-S without finding materials which gave synthetic rubber the tack of natural rubber. All measurements of tack were based on hand tests.

The report of the committee investigating the German Rubber Industry¹¹ indicates that one of the few worthwhile advances made by the German Rubber Industry during the war was the development of Koresin as a tackifier of butadiene-styrene rubber. This was reported to give the rubber excellent tack, but no quantitative data were obtained.

III. ANALYSIS OF "TACKINESS"

Before trying to measure quantitatively the tackiness of synthetic rubbers, it is helpful to analyze in detail the operation of measuring tack, to determine some of the different kinds of factors which are involved. This operational method of analyzing the process leads to a recapitulation and organization of many of the factors which have been suggested in the literature and to the recognition of some additional factors.

Expressed in the simplest terms, the operation of determining tackiness consists of:

- (A) *Producing contact between two surfaces for a short time by the application of a force, and*
- (B) *Separating the surfaces and noting the force required.*

The force applied to separate the surfaces is the "tackiness" under the particular test conditions used. Since most tack tests have been made by hand, this implies that the magnitude of the adhesion force on the area tested (usually of the order of 1–10 sq. cm) should be great enough to be detected by the tactual sensations of the hand. Mechanical tests, however, can extend the concept to a much wider range of values. Further consideration shows that each of these operations can be broken down into a number of simpler factors, as follows.

⁶ D. Josefowitz and H. Mark, *Ind. Rubber World* 106, 33 (1942).

⁷ Zhukov and Talmud, *J. Rubber Ind. U.S.S.R.* 12, 1005 (1935).

⁸ *Rheology Bulletin* 16, 70 (1945); *Paint Technology* 9, 211–221 (1944).

⁹ J. R. Scott, *Paint Technology* 9, 218 (1944).

¹⁰ L. E. Ludwig, D. V. Sarback, B. S. Garvey, and A. E. Juve, *Ind. Rubber World* 111, 180 (1944), and 112, 731 (1945).

¹¹ Summary Report, Rubber Bureau WPB, and Office of Rubber Reserve, Washington, D. C., August, 1945.

A. Producing contact between two surfaces.

The amount of surface brought into "molecular contact" in a given area depends on:

1. Initial roughness of the surface. This determines the magnitude of the deformations required to bring the surfaces into "contact," i.e., within the range of the molecular forces in the surface, or within say about 10^{-8} cm.

2. Forces acting to bring the surfaces together. These include:

- (a) External forces applied to the material. These usually range from about 1 g to 1 kg/cm² of nominal area and they may act for times ranging roughly from a few seconds to a minute. Occasionally we are interested in the behavior of a material when surfaces are brought together under pressure for times up to weeks or months, as when bales or sheets of a plastic are piled together in storage. Sealing of the surfaces under these conditions is sometimes called "blocking."

- (b) Molecular forces at the surface of the material. These short range forces come into play only when the surfaces are in very close contact. They are usually greatest for polar materials, as indicated by the fact that such materials have the highest surface tension in the liquid state. While many of the most tacky materials are polar (asphalt, (warm) rosin, paint, polyvinyl ethers, etc.), the presence of polar molecules is neither necessary nor sufficient for tackiness. Some materials such as polyisobutylene are not polar but are very tacky while rosin in the solid state is hard and brittle, rather than tacky. In multi-component systems the tackiness may be determined by the molecular forces of the component that diffuses or blooms to the surface in the highest concentration.

3. Resistance of the material to the deformation required to bring the two surfaces into molecular contact. This involves the visco-elastic properties of the material for deformations of both:

- (a) a relatively large scale, required to overcome the gross surface roughness of the material, and
- (b) a very small scale, required to produce molecular contact between even smooth surfaces. Since the molecular forces have such short range, the presence of even a small amount of dust may prevent large fractions of the surfaces from

coming close enough to be within the range of their cohesive force unless these small scale deformations can be made very easily.

It should be noted that the modulus, or the resistance to deformation, for these very small scale deformations at the surface may be very much higher or very much lower than the modulus for large deformations of the whole sample. A very thin layer of hard oxidized material might greatly increase the modulus of the surface layer, while the blooming of a small amount of liquid component to the surface might reduce the modulus of this layer. It is possible that the presence of small regions of rigid gel structure at the surface of GR-S might tend to prevent the molecular contact needed to produce good tack.

The relaxation times of material in the surface layer and the bulk material determine the length of time the forces must act to produce the required deformations.

B. Separating the surfaces and noting the force required.

This factor will depend on:

1. The magnitude of the molecular forces per unit area of molecular contact, and on the total area of such contact.

2. The ultimate tensile strength of the material for the particular kind and rate of applied stress.

It is this factor which keeps liquids such as water from being tacky. The water is easily deformed, makes good molecular contact with other surfaces brought into contact with it, and it has high surface forces, but it does not have sufficient tensile strength or viscosity to resist separation, even at high speeds. In the case of viscous liquid materials like tars, printing inks, etc., the "tensile strength" may in effect be merely the viscous resistance to deformation. Here the "tensile strength" depends very much on the rate of deformation.

3. The distribution of the applied stress and resultant elongations. This determines the magnitude of the force which must be applied to stress the material to its "tensile strength" at the line of separation. This depends on:

- (a) The geometry of the sample, i.e., the thickness, area, and shape, and the position and direction of application of the applied force.

This may be applied normal or parallel to the surface, or to produce peeling apart of the materials along a line, producing various kinds of stress gradients.

Since in tack tests the separating force is applied in such a way as to produce a stress gradient in the sample, the total force will depend very much on the ultimate elongation of the material, for this affects the total volume of material that must be stretched.

It should be noted that under some conditions, the force applied to remove the samples may result in two- or even three-dimensional tensile stresses in the sample. The behavior of the sample to such stresses may be quite different from its behavior under a one-dimensional tensile stress, just as is the case in tensile and tear tests of cured rubber.¹² As a rule, the most tacky materials react to this complex stress pattern by forming irregular "teeth" or "legs" in the region of separation.

(b) The stress-strain-relaxation properties of the material.

General experience suggests that the differences in apparent tack which are found by varying the conditions under which the surfaces are separated (e.g., at different rates of separation, temperature, etc.) probably are due to the changes in the stress distribution in the region of separation, resulting from the differences in the apparent effective visco-elastic properties of the medium.

Two sheets of milled rubber, for example, which are stuck together so firmly that they tear along new surfaces when pulled apart slowly, can sometimes be separated along the original surfaces by a very quick pull. This in effect makes the rubber much stiffer and stronger, and thus puts relatively more stress on the bond between the surfaces.

IV. MEASUREMENT OF TACKINESS

(1) Effect of Testing Conditions on Tack of Adhesive Tapes

The properties which affect the tack (or resistance to separation) such as the modulus, elongation, and plastic deformation, vary with both time of application of force, and temperature of the material. Therefore, a complete description

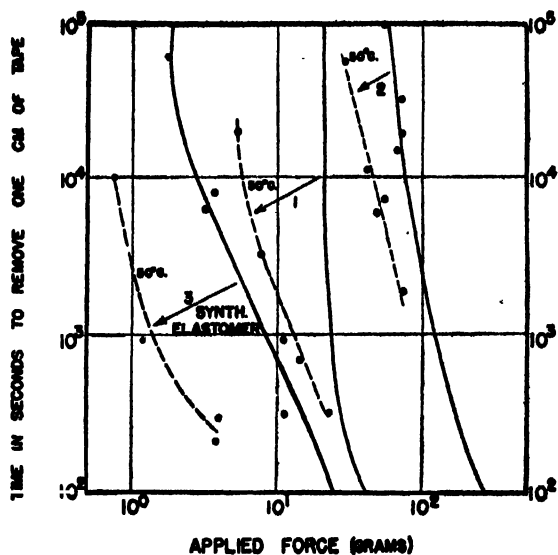


FIG. 1. Relation of applied force to rates of removal of adhesive tapes.

of the tackiness of a material should include data for a wide range of time of contact of load and rate of separation of surfaces, preferably also over a wide range of temperatures.

To get some quantitative data on the importance of the rate factor, measurements were made of the relation between the applied load and the rate of removal of pressure sensitive adhesive tapes from glass surfaces, since this makes the factor of surface roughness small and constant.

The simplest method used to measure this relation was to put a number of strips of adhesive tape $\frac{1}{4}$ in. wide by about 12 in. long (1.25 cm $\times$ 30 cm) on the bottom of a 10 in. by 15 in. (25 cm $\times$ 37 cm) plate of glass held horizontally in a rack, hang known loads from the free ends of each tape to pull it away from the glass. This method had the advantage that it could be used in an oven to get data at elevated temperatures, and a number of tapes and loads could be tested at any one time. It had the disadvantage that, particularly at low loads, non-uniformities in the tape might make the rate of removal irregular, or even stop it altogether. However, the test gives considerable more information about the adhesive properties than a test which measures only the force required to remove the tape at one rate.

Data obtained with this test for three samples of experimental pressure-sensitive adhesive tapes

¹² W. F. Busse, J. App. Phys. 9, 438 (1938).

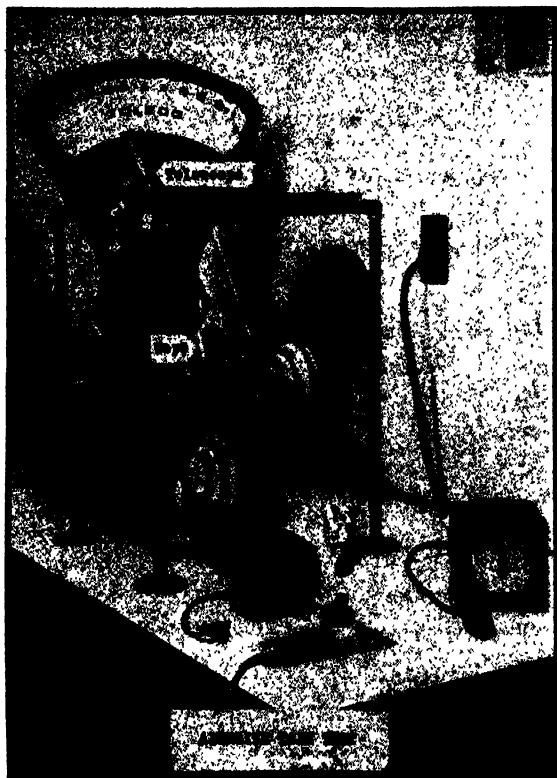


FIG. 2. Adhesive tape test.

at room temperature and 50°C are shown in Fig. 1. Tape No. 1 is made using GR-S, No. 2 has natural rubber, and No. 3 has an experimental synthetic elastomer base.¹³ Since each tape has various amounts (not necessarily the optimum) of plasticizer and tackifier, the results illustrate the physical properties that can be obtained in tapes, rather than the fundamental properties of specific elastomers. Because of the long times involved, the removal of the tapes are expressed as reciprocal rates, or seconds/cm.

At room temperature tape No. 3 will support a load of 1.8 g indefinitely (10^5 sec. to remove 1 cm) but it requires about 27 g or a 15-fold increase to remove it at the rate of 10^2 sec./cm. The tape made with natural rubber will hold a load of 20 g indefinitely, but 40 g will remove the tape at 10^2 sec./cm, or a range of only two-fold, which is very desirable. The GR-S tape will support larger loads, but about a fivefold range of loads is required to change the reciprocal rate of removal from 10^5 sec./cm to 10^2 sec./cm.

¹³ These were supplied through the courtesy of Industrial Tape Corporation, New Brunswick, New Jersey.

The loads required to remove the tapes at a given rate at 50°C are seen to be only about one-half to one-fifth those required at room temperature, but not enough data are available to draw any theoretical conclusions from the temperature coefficient.

To avoid some of the difficulties of this first test, equipment was built by Mr. F. E. Wetherill of this laboratory to remove the tapes at definite rates over nearly a 10^6 -fold range and measure the force applied. This was done by fastening the glass plate to the pan of a direct reading scale that read from 1 to 1000 g. The adhesive tape, 1.2 cm (0.5 in.) wide by about 25 cm (10 in.) long, was fastened to the top of the glass plate, and the free end of the tape was pulled normal to the surface by means of a cord which passed over a roller about a foot above the tape, and then went around another pulley which was driven at the desired speed.

The wide range of driving speeds which were required were obtained very simply by a train of multiple pulleys made from wood and fiber board. They were mounted so that, for example, the large pulley of each unit could be driven by friction from a smaller pulley of the preceding unit. By changing the number and diameter of the pulleys used in the train, and using a motor with built-in speed reducer, the speed of removal of the tape could be varied from about 10^{-1} to 5×10^4 sec./cm.

The amount of tape removed in a given time interval was read with a small telescope from a metal scale placed on the balance pan beside the tape. A picture of the equipment is shown in Fig. 2.

Typical results from this test are shown in Fig. 3 which gives data on experimental tapes. The elastomers are No. 1, natural rubber, No. 2, GR-S, No. 3, and No. 4, two other synthetic elastomers. Again the data illustrate only the behavior of the particular tapes rather than fundamental properties of the elastomers.

It is apparent that tapes Nos. 1, 3, and 4 have about the same properties if tested at relatively high rates of removal—say, from 0.1 to 5 sec./cm removed (10 to 0.2 cm/sec.). At very low rates of removal, however, they are quite different. Tapes Nos. 3 and 4 are removed slowly at loads above about 3.3 g, while the rubber base tape

does not start to be removed until the load is over 10 g. The behavior at these low rates of removal is important wherever pressure-sensitive tapes are used to seal packages, or to hold objects on a wall, etc., as they are then subject to small stresses for long times, and their performance in service may be determined by this factor. Tapes such as No. 3 and No. 4 failed on such service tests while No. 1 was satisfactory. The tape No. 2 had good adhesion to the glass, but other properties made it less satisfactory than No. 1.

(2) Tackiness Test for Natural Rubber and GR-S

a. Discussion of Test

The test used in this work was designed primarily to give a quantitative measure of the effect of various materials on the tackiness or resistance to separation of sheets of GR-S under conditions approximating those of the hand test. The method could easily be modified to study the effects of varying the magnitude and time of

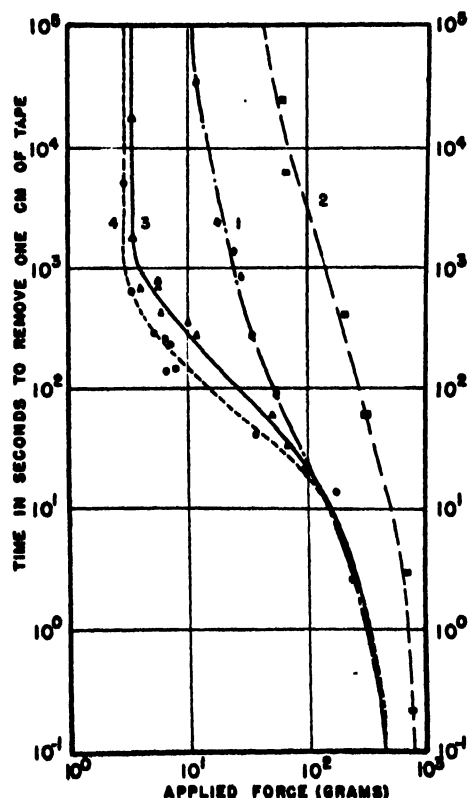


FIG. 3. Relation of applied force to rates of removal of adhesive tapes.

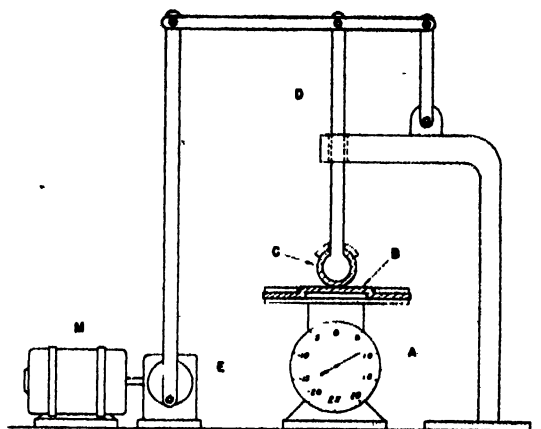


FIG. 4. Schematic diagram of tackiness tester.

application of the compression force, the rate of separation of the surfaces, etc. However, this was not found necessary for the evaluation of compounding ingredients and milling conditions on the tackiness of GR-S.

An obvious difficulty in developing a quantitative test of the tackiness of rubber is to get samples with reproducible smooth, uniform surfaces. A number of ways of doing this have been suggested, including molding smooth samples, depositing the rubber from solution, calendering the rubber onto fabric, etc., but they all have the disadvantage of changing the rubber that is to be tested.

A closer examination of the problem shows, however, that surface roughness is not necessarily as serious a problem as it first appears. Since it has been possible in practice to ignore this factor in hand tests, it is evident that by standardized preparation of the sample, surfaces of at least approximately reproducible roughness can be obtained. It was found possible to use rubber samples 2-3 mm thick prepared by milling sheets on a small (2"×6") laboratory mill (made by Wm. R. Thropp and Sons). Uncompounded GR-S gave the most irregular samples, but on adding 45 parts EPC black to 100 parts GR-S the magnitude of the irregularities was reduced, and the sheets appeared more uniform. However, even the uncompounded GR-S could be prepared in fairly reproducible sheets.

To measure the tackiness, the equipment shown schematically in Fig. 4 was used. One strip of the rubber to be tested, B, at least

TABLE I. Effect of milling on tack of pale crepe.

Sample	Treatment	Temp. °C	Time, min.	Separating force-g after		Appearance of sample after milling
				1-2 hr.	24 hr.	
I	Initial milling	60	8	525	185	Smooth
Ia	I remilled	25	7	115	—	Creped-rough
Ib	I remilled	60	7	405	—	Smooth
II	Initial milling	25	3	30	35	Creped-rough
IIa	II remilled	25	3	35	—	Creped-rough
IIb	II remilled	60	3	330	—	Smooth
III	Initial milling	25	15	105	110	Creped-rough
IIIa	III remilled	25	15	600	—	Smooth

3×5 cm (1.2×2 in.) is fastened to the pan of a pre-loaded spring scale, *A*, which reads compression and tension loads directly. Another sample, *C*, of the rubber is put around a mandrel 3.4 cm (1½ in.) in diameter and 2.5 cm (1.0 in.) wide and lowered by the motor *M* through the drive *B*, in a 2.4-sec. cycle.

The sample *C* is pressed against *B* with a force of 565 g (20 oz.) at the bottom of the stroke and then pulled away. The maximum position of the scale pointer on the return stroke is read. This value, when corrected for the inertia effects of the system, is a measure of the force required to separate the samples, or the "tackiness," for "instantaneous" loading. At least five tests are made at different places on the surface and averaged for the "instantaneous tack" values.

The motor driving the mandrel can be stopped at the bottom of the cycle to leave the samples in contact under load for any desired time, and the motor then re-started. The force required to separate the samples after the load is applied for 10 sec. is called the "10-sec. tack." Five readings of this are taken and averaged.

Usually the "instantaneous tack" and the "10 sec. tack" values changed more or less parallel with each other, so the two values are averaged in this report unless otherwise noted. In a few cases, where the samples sealed and could not be separated after the load was on for 10 sec., only the instantaneous values are given. The average tack values for any one sample are reproducible to within about ±30 grams.

TABLE II. Effect of milling temperature on relative tack of GR-S.

Temperature — °C	15	30	50	70
Separating force (g)	155	85	30	10

Because the initial rubber and the details of processing were different in different series of tests, the absolute values of tackiness in different series cannot be compared. Within each series, however, the data are comparable.

b. Effect of Some Processing Conditions on Tack of Natural Rubber and GR-S

It is known that natural rubber breaks down (due to oxidation) and becomes soft and plastic more rapidly when milled on a cold mill (20°C) than when milled on a warm or hot mill¹⁴ (100°C) so it might be expected that cold milling would make the rubber more tacky than hot

TABLE III. Effect of mastication time on tackiness and plasticity of GR-S.

Composition	Mast. time, min.	Separating force (g) when mast.		Williams' plasticity (100°C) when mast.	
		Hot	Cold	Hot	Cold
GR-S	5	215	160	2.9	3.0
	10	285	200	2.7	2.8
	20	425	285	2.3	2.7
	30	525	300	2.1	2.5
	45	640	315	1.8	2.4
GR-S and EPC black	20	115	55	4.4*	4.1
	30	115	55	4.2*	3.8
	45	140	85	3.9*	3.7
	60	170	85	3.8*	3.5

* Plasticity values after 9 days.

milling. The following test (Table I) shows that this result is not always obtained.

Samples of pale crepe were milled at 25° and 60° on a 2"×6" mill, the temperatures of the rubber being measured by a thermocouple. The tack values or separating force (grams) were measured within 1-2 hr. and after 24 hr. Samples given the initial hot milling were then remilled at 60° and at 25°C. Some of the samples were so tacky that the surfaces sealed when pressed together for 10 sec., so all of the tack values of Table I are for the separating force in grams after the "instantaneous" load.

The results shown in Table I agreed qualitatively with the hand tests of the tack of these samples. They suggest that the difference in tack of the hot- and cold-milled samples is largely due to the physical state of the surface. The samples

¹⁴ W. F. Busse, Ind. Eng. Chem. 24, 140 (1932); Proc. Rubber Tech. Conf. London, May 1938, p. 288; Cotton, Inst. Rubber Ind. Trans. 6, 487 (1931).

which were cold milled for a short time had rough creped surfaces which prevented intimate contact over large areas, so gave low values for the tackiness. The hot-milled samples, and the samples cold milled for 30 minutes had smooth surfaces which allowed good contact and so gave very high tack values. The decrease in tack of sample 1 on standing 24 hr. may be due in part, at least, to the development of surface roughness on standing.

The results show that the tackiness measured in this test is not an absolute property of the material. However, it does appear to measure a combination of properties of the material and the sample geometry which is important both in the hand tack test, and in "adhesion" of plies in actual factory processing.

When GR-S was milled on the small mill the tack was low for all milling temperatures from 15 to 70°C, as shown in Table II. The tack is somewhat higher for samples milled cold, but varying the milling time from 5 to 15 min. had little if any effect on the tack. Samples of GR-S from different sources gave some variation in tack values, but none were very tacky.

When GR-S and a stock containing GR-S 100 parts and EPC black 45 parts were masticated for various times in a Midget Laboratory Banbury mixer the results shown in Table III and Fig. 5 were obtained. The samples marked "hot" were run with steam (100°C) through the jacket and cold water through the rotor, and

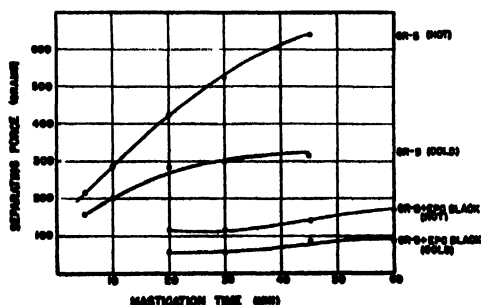


FIG. 5. Effect of mastication time on tackiness of GR-S.

those marked "cold" were run with cold water (20°C) circulating through the jacket and rotor. 250 g of rubber were used and 30-g samples were removed for test after various times of mastication. For the stock containing carbon black the rubber was broken down for 5 minutes and the

TABLE IV. Effect of polyvinyl isopropyl ethers on tackiness of GR-S.

Sample	Specific viscosity*	Separating force (g)
1	0.05	50
2	0.13	160
3	0.63	330
4	0.84	310
5	2.13	280
Control (GR-S only)	—	30

* In 1 percent benzene solution.

carbon black was then added and mixed 15 minutes when the 20 minute sample was removed. Each sample was then milled 2 minutes on the 2"×6" mill to sheet out, the rubber samples being milled at 20°C and the black samples at 600°C. Tack tests and Williams plasticity tests at 100°C were made after 24 hr., except for one group of plasticity tests made after standing for a longer time. The plasticity figures are the thickness in mm of the 2-cc sample after being subject to the 5 kg load for 3 min. The tack values are the average of "instantaneous" and "10-second" tests for each of two independent series of mastication tests.

It is seen that in these tests the hot Banbury mastication consistently gives somewhat tackier and more plastic samples than the cold mastication, even with the stock containing 45 parts EPC black. Increased mastication time increases the tack readings, in part because it makes the samples softer. This method of increasing tackiness may not be practical, however, because of the excessive time of mastication required and the deleterious effects of high temperature on the properties of the cured compounds.¹⁵

c. Effect of Vinyl Ethers on Tackiness of GR-S

Since some of the polyvinyl ethers are extremely tacky materials themselves, it was thought that they might improve the tackiness of GR-S. A series of polyvinyl ethers¹⁶ of different molecular weights (viscosities) was tested by milling 10 parts each into 100 parts GR-S on the 2"×6"-mill run hot (60°C). The results are shown in Table IV.

¹⁵ W. B. Wiegand and H. A. Braendle, *Ind. Eng. Chem.* **36**, 699 (1944).

¹⁶ Prepared by Drs. F. Grosser and C. E. Schildknecht of this laboratory.

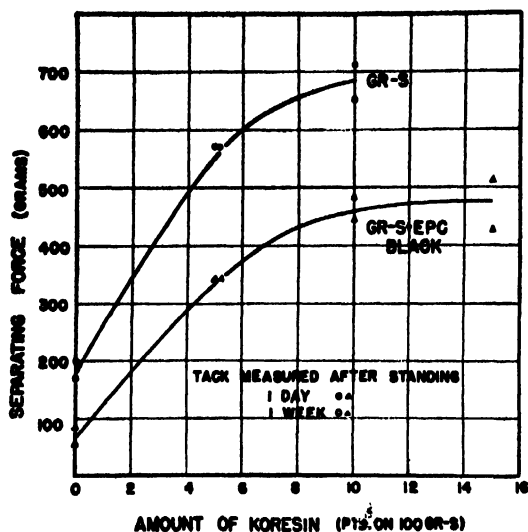


FIG. 6. Effect of mixing Koresin into GR-S.

The results show that within the proper molecular weight range, these vinyl ethers do produce tackiness in GR-S but the lower molecular weight ethers were not effective. On the same test, 10 parts of rosin oil in 100 parts GR-S did not produce a measurable increase in the tackiness.

The vinyl ethers are not very compatible with the GR-S rubber, so they probably tend to bloom to the surface and thus make the surface tacky. Compounded stocks containing vinyl ethers can be plied up and cured satisfactorily to a solid piece. However, if two samples are *painted* with a solution of vinyl ether, and pressed together during cure, they will not fuse together, and the cured samples can easily be separated.

d. Effect of Koresin and Other Tackifiers on GR-S

When reports reached this country that Germany had developed a product that was a very effective tackifier of their Buna S, the Rubber Reserve took steps to find its composition and to get a source of supply in this country. The material was found to be a reaction product of acetylene and *p. ter* butylphenol. Rubber Reserve made arrangements to have this product, known as Koresin, made in this country¹⁷ and samples supplied to the rubber industry.

¹⁷ By General Aniline and Film Corporation, USP 2,337,464.

The results obtained when various amounts of Koresin were mixed in GR-S and in a stock containing GR-S 100 parts and EPC black 45 parts are shown in Table V and Fig. 6. The mixing was done for 5 minutes in the small Midget Laboratory Mixer run hot, and the samples were then sheeted out for 5 minutes by milling on a warm (60°C) 2"×6" mill. The tackiness was measured after the samples stood various times up to a week in an air-conditioned room 75°F-30 percent R.H.

It is seen that 5-15 parts Koresin on 100 GR-S are very effective in increasing the tack of both straight GR-S and of GR-S containing 45 parts black and that the tack decreases only slightly on standing for one week. However, if the samples are stored at high humidities or where water gets on the surface, the tack drops more rapidly. The addition of the Koresin has no deleterious effect on rate of cure, or tensile properties of the cured GR-S compounds.

When Koresin is applied as a solution or cement to the surface of a stock containing GR-S 100 parts and EPC black 45 parts, the results depend very much on the concentration used. Table VI and Fig. 7 show the results of

TABLE V. Effect of mixing Koresin into GR-S.

Stock	Parts Koresin	Separating force (g) after standing for		
		1-2 hr.	24 hr.	1 week
GR-S	0	200	170	200
	5	540	570	570
	10	650	710	650
GR-S and EPC black	0	55	55	85
	5	395	340	340
	10	455	480	455
	15	595	425	510

two series of tests—one studying the effect of various concentrations of Koresin solution in hexane, and the other comparing Koresin with other materials when applied as solutions in hexane.

There is a definite optimum at around 3-4 percent concentrations of Koresin solution. This tends to confirm the view that the Koresin increases the tack of GR-S in part, at least, because of its tendency to bloom to the surface, forming a layer of high Koresin concentration. When a large excess of Koresin is present, how-

ever, the Koresin seems to form a "lacquer" on the surface which is dry, hard, and non-tacky.

If the solutions are applied in atmospheres of very high humidity, where moisture can condense on the surface, the tackiness is very much reduced. The mechanism of this action is not known, but a similar effect is noted with rubber cements.

A comparison of a number of different softeners and tackifiers added in the Midget Banbury Mixer to a stock containing GR-S 100 parts and EPC black 45 parts is shown in Table VII. The plasticity values are taken after 24 hr., on the Williams Plastometer at 100°C.

It is seen that 10 parts Koresin gives the most tack of any material tested. A mixture of equal parts of Koresin plus the petroleum derivative (Circosol 2XH) (10 parts total) is almost as effective as 10 parts Koresin, even though 10 parts of the petroleum derivative alone has little effect on the tackiness. The tackiness of the Koresin stocks does not decrease markedly in a week when stored in an atmosphere at 75°F and 30-35 percent relative humidity.

CONCLUSIONS

The analysis of tackiness shows that it is a complex property which depends on a balance between a large number of different factors, including the geometry of the sample as well as the surface forces and the visco-elastic properties of the material itself.

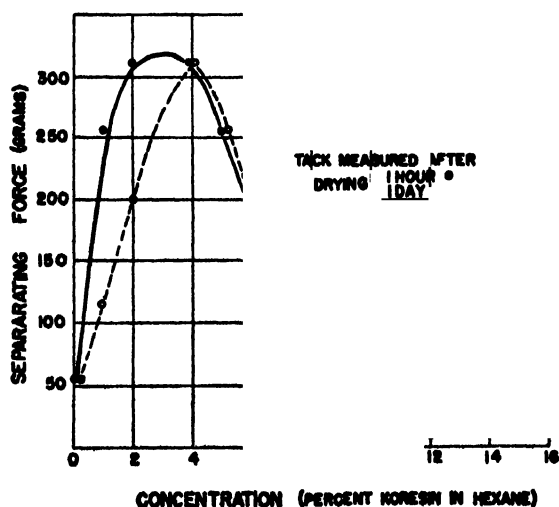


FIG. 7. Effect of painting Koresin solutions (Hexane) on surface of GR-S carbon black compound.

TABLE VI. Effect of painting various materials on surface of GR-S carbon black compound.

Material	Conc. %	Separating force (g) after			
		1 hr.	3 hr.	24 hr.	1 week
<i>Series 1.</i>					
Solvent only	—	55	—	55	—
Koresin	1.0	255	—	115	—
Koresin	2.0	310	—	200	—
Koresin	4.0	310	—	310	—
Koresin	5.0	255	—	255	—
Koresin	10.0	30	—	30	—
Koresin	15.0	30	—	30	—
<i>Series 2.</i>					
Koresin	3	255	225	255	140
Koresin (German)	4	225	170	140	140
GR-S	4	115	85	85	85
Natural rubber	4	370	225	170	55
Dehydro abietic acid	4	170	140	85	55

TABLE VII. Effect of various materials on the tackiness of GR-S EPC black compound.

Material added	Amount added (on 100 pts. GR-S)	Separating force (g) after standing for			Williams plasticity 100°C
		1-2 hr.	1 day	1 wk.	
Control (GR-S comp. only)	—	55	55	55	4.3
Koresin	5	395	340	340	3.9
Koresin	10	455	480	455	3.6
Koresin and petroleum deriv. (1:1)	10	480	455	455	3.7
Koresin and coal tar fractions (1:1)	10	425	340	425	3.7
Dehydrogenated rosin	10	370	255	300	3.7
Petroleum derivative	10	170	85	115	4.1
Coal tar fraction	10	170	115	85	3.8
Rosin oil	10	300	115	115	3.7
Stearic acid	10	55	80	55	3.8

The quantitative measurements of the effect of the rate of removal on the apparent tackiness of pressure-sensitive adhesive tapes show that significant differences are found at very slow removal rates (10⁵ sec./cm) which do not show up at much higher removal rates (10 sec./cm). These differences may affect the performance of the tapes in service.

A simple quantitative test is described which can be used to study the effects of various processing treatments and various softeners, tackifiers and other compounding ingredients on the tackiness of GR-S and other rubbers. Koresin was the most effective tackifier found for GR-S.

The authors wish to acknowledge the continued interest of Dr. W. E. Hanford, Director of the Laboratory, and the helpful discussions of tackiness measurements with Dr. C. E. Schildknecht. Miss Ruth Sellers and Miss J. L. Nuver helped in taking much of the data. The first Koresin samples were prepared in the laboratory by Drs. C. E. Schildknecht and A. O. Zoss before the commercial material was available.

Applications of Infra-Red Methods in the Structural Examination of Synthetic Rubber

J. E. FIELD, D. E. WOODFORD, AND S. D. GEHMAN

Research Laboratories, The Goodyear Tire & Rubber Company, Akron, Ohio

The physical properties of a synthetic polymer are associated with variations in molecular structure which are dependent on the conditions of the polymerizing reaction and the monomers used. In some cases, structural differences of significance for physical properties cannot be detected by infra-red methods because the relative number of chemical linkages affected is too small. An important structural detail which can be followed by infra-red analysis is the relative amount of 1,2 and 1,4 polymerization occurring in polymerization reactions of butadiene. Absorption curves are reproduced to show the wide range in the relative amounts of these two structures. A description is given of an attempt to obtain a quantitative measure of this ratio by means of a calibration curve derived from known mixtures of pure octene 1 and octene 2. For polyisoprene, variations in the proportions of 1,4 and 1,2 or 3,4 structure also occur depending upon the type of polymerization. Comparison with Hevea rubber and Balata fails to disclose definite evidence of *trans*-isomerism in synthetic polyisoprene. Structural differences due to oxidation of polymers may be readily apparent in infra-red spectra, hydroxyl and carbonyl groups being especially prominent. The effectiveness of anti-oxidant in preventing structural changes caused by oxidation is shown in a series of absorption curves for samples of GR-S with and without anti-oxidant. When the samples were heat-treated in air, pronounced structural changes occurred for the sample without anti-oxidant, but no perceptible changes were evident for the sample with anti-oxidant. On the other hand, the anti-oxidant used (phenyl- β -naphthylamine) was ineffective for stabilizing the structure toward ultraviolet light.

INTRODUCTION

INFRA-RED absorption spectra have been long recognized as a convenient means for studying the structure of organic molecules. The interpretations of the spectra are based on the energy interactions of the molecule and the radiations which arise from the vibration of the constituent atoms and molecular rotations. For simple or highly symmetrical molecules, the determination of the normal modes of vibration and the calculation of the absorbing frequencies are relatively simple and straightforward. For more complicated organic molecules, this becomes increasingly difficult because with each additional atom, the number of degrees of freedom is increased by three and the determination of the normal modes of vibration becomes practically impossible. However, interpretations can be made to a useful extent through empirical comparison with the absorption spectra of simpler known structures.

The data that have been accumulated by investigators in this field have made it possible to assign rather definite absorption frequencies to some of the chemical linkages and functional groups. These correlations which have appeared

in numerous places in the literature are partially reproduced in Table I.

Organic compounds generally have strong absorption bands below 1300 cm^{-1} to which few definite assignments can be made with certainty because the vibrations of many of the atoms of the molecule may be involved rather than a specific part of it. It is quite clear that such empirical relationships must be relied upon in studying the structural variations of the long chain, complex molecules which occur in butadiene and isoprene polymers and copolymers and other synthetic rubbers. This procedure has been applied to determine the effects of oxidation and of variations in monomers and polymerizing conditions on the structure of synthetic rubber. It is practically certain that physical deficiencies of synthetic rubber are due principally to the structure of the long chain molecules rather than to the chemical nature of the monomers used.

TABLE I.

Frequency range	Chemical linkage
3000-4000 cm^{-1}	CH, OH, NH
2000-2500 cm^{-1}	C $\equiv$ C, C $\equiv$ N
1350-2000 cm^{-1}	CH ₂ , CH ₃ , CH, C=C, C=O
900-1000 cm^{-1}	C=C

EXPERIMENTAL PROCEDURE

A conventional Littrow type rocksalt prism spectrometer of our own design and construction, shown in Fig. 1, was used for obtaining the absorption records. The refracting faces of the 60-degree prism are 15 cm long by 10 cm high. The current from the vacuum thermocouple is conducted to a galvanometer the deflections of which actuate a photo-cell amplifier. The absorption curves are automatically recorded by a Leeds and Northrup Speedomax which operates from the amplifier. These curves show the absorptions (not percent absorption) by the varying galvanometer deflection recorded as a function of the wave number setting of the Littrow mirror.

The samples were prepared, in most instances, by evaporating the solvent from polymer solutions poured on rocksalt plates, thus leaving a thin film of polymer for examination. The thickness of the film was adjusted for about 10 percent transmission of the 1450 cm^{-1} band. This was rather easily accomplished by controlling the concentration of the solution and by the use of successive layers of film. The latex samples were prepared by placing a few drops in a closed glass frame floating on a mercury surface. In order to prevent undue oxidation, these films were first dried in a nitrogen atmosphere and then in a vacuum desiccator. Since the film thickness of the various samples was maintained essentially the same, a fair estimate of the intensity of the absorption of a particular band can be made relative to a corresponding one of another sample



FIG. 1. Spectrometer and recorder.

without transforming the absorption curves into terms of percent transmission.

ABSORPTION SPECTRA OF VARIOUS POLYMER CHAIN STRUCTURES

The chemistry of polymerizing reactions is rather complex in character. Any discussion concerning the general principles and mechanisms of such reactions need not be entered into here. However, in order to understand the absorption spectra of polymers of butadiene and isoprene, which have played such an important role in the development of synthetic rubber, it is necessary to point out the various ways in which polymerization of these diolefinic monomers can proceed. These are diagrammatically shown in Fig. 2. The structures appearing at the double

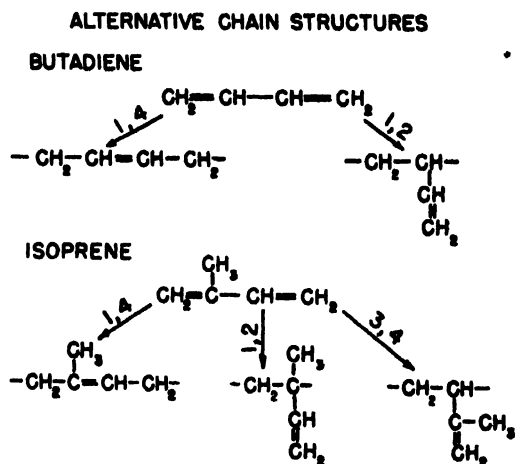


FIG. 2. Alternative chain structures for polybutadiene and polyisoprene.

bonds may be typified in the manner shown in Table II. Compounds which can be differentiated according to the above classification are known to have characteristic absorption frequencies which will be of considerable value in the interpretation of the absorption spectra to be discussed.

TABLE II.

Polybutadiene 1,4 addition 1,2 addition	$\text{R}_1\text{CH}=\text{CHR}_2$ $\text{RCH}=\text{CH}_2$
Polyisoprene 1,4 addition 1,2 addition 3,4 addition	$\text{R}_1\text{R}_2\text{C}=\text{CHR}_3$ $\text{RCH}=\text{CH}_2$ $\text{R}_1\text{R}_2\text{C}=\text{CH}_2$

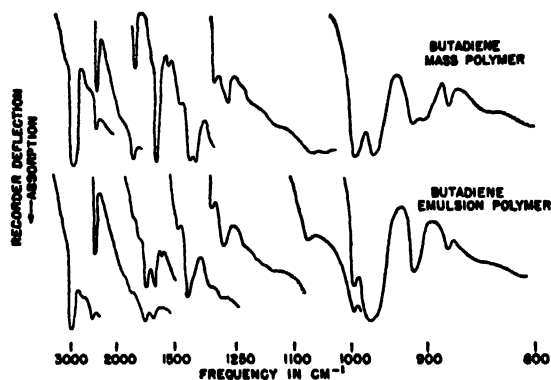


FIG. 3. Absorption curves for sodium and emulsion polymers of butadiene.

POLYBUTADIENE AND GR-S

The absorption curves for sodium and emulsion polymers of butadiene are shown in Fig. 3. The differences appearing in the two spectra are quite significant. That a high degree of 1,2 polymerization occurs in the sodium polymer has been previously deduced from ozonolysis and infra-red work.¹⁻⁴ This is recognized by the appearance of the very strong band at 996 cm^{-1} which is characteristic of compounds of the type $\text{RCH}=\text{CH}_2$.^{2,3,5} A relatively weak band appears at this frequency for the emulsion polymer. Absorptions are present in both polymers at 967 cm^{-1} which is characteristic of compounds of the type $\text{R}_1\text{CH}=\text{CHR}_2$. The bands at 914 cm^{-1} and 1450 cm^{-1} appear to be split into doublets in the sodium polymer, evidently because of the influence of the large number of terminal double bonds.

In the development of synthetic rubber of the GR-S type, it was found that emulsion polymers possessed the more desirable physical properties and processing qualities. Since the product of a reaction of this type for butadiene leads predominately to the 1,4 structure, it is thought that the relative amounts of these two structures

in the copolymer has some bearing on these physical characteristics because butadiene is the principal constituent in the GR-S type of synthetic. Accordingly, methods have been sought to obtain quantitative determinations of relative amounts of 1,2 and 1,4 structure in polymers and the GR-S copolymer. Infra-red methods afford a positive and direct way of carrying out such analyses. The following procedure has been found to be convenient for this purpose. Referring to Fig. 4, which illustrates the resultant band due to the two overlapping bands at 996 and 967 cm^{-1} , it is evident that the ratio A/B can be used for characterizing the band structure. If this ratio can be calibrated in terms of known relative amounts of the two structures, the infra-red absorption spectrum immediately becomes applicable as a method of determination. In order to do this, absorption spectra were obtained for two compounds, one having a terminal double bond and the other an internal double bond. Pure samples of octene-1 and octene-2 were used for this purpose. The spectra of known mixtures of these hydrocarbons were obtained and the ratio A/B was determined from the absorption curves. This ratio was then plotted as a function of percent octene-1 shown in Fig. 4. It is then assumed that this curve applies for the butadiene polymers and copolymers. This assumption may not be entirely valid since other structural effects on these bands are not considered, but it seems reasonable to believe that a very close approximation can be reached with this method. Following the procedure outlined above, the results for several typical polymers are given in Table III.

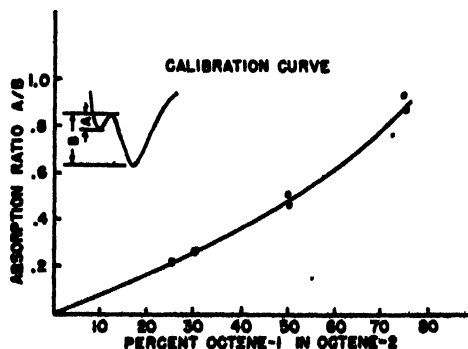


FIG. 4. Calibration curve for determining the amount of 1,2 structure in polymers and copolymers of butadiene.

¹ N. Alekseeva, *Rubber Chem. Tech.* 15, 693, 698 (1942).

² Unpublished work of M. W. Swaney, J. U. White, and J. E. Flory, Esso Laboratories Standard Oil Development Company.

³ H. W. Thompson and P. Torkington, *Trans. Faraday Soc.* 41, No. 279, 246 (1945).

⁴ Unpublished work of R. S. Rasmussen and R. R. Brattain, Shell Development Company.

⁵ M. Tuot, J. Becomte, and S. Lorillard, *Comptes rendus* 241, 586 (1940).

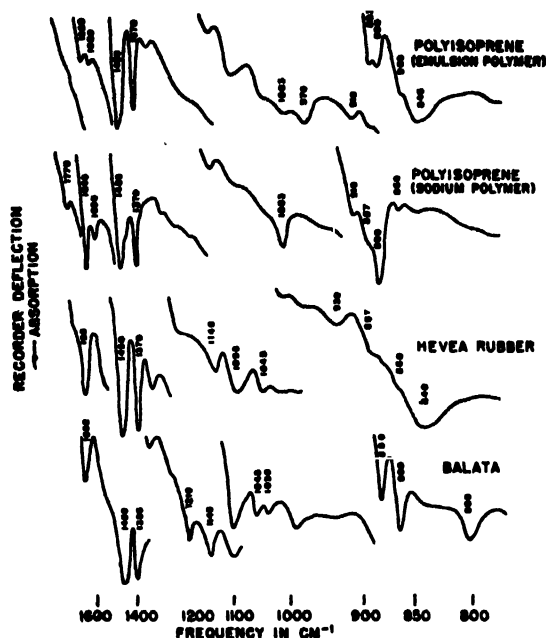


FIG. 5. Absorption curves for comparing the structure of Hevea rubber, Balata, and the emulsion and sodium polyisoprenes.

The GR-S sample included in the above table was a typical production sample. It is of interest because it shows that the amount of 1,2 structure determined by this method is within reasonable agreement with estimates obtained by other methods. The extremely high percentage shown for the sodium polymer may not be so accurate since the method appears to be less sensitive for mixtures with such high ratios.

POLYISOPRENE

The absorption spectra for the emulsion and sodium polymers of isoprene are shown in Fig. 5. The structural study of this polymer is of particular interest because the monomer unit is the same as for natural rubber and yet the structure and physical properties of the latter cannot be reproduced in the synthetic product. The absorption spectra for Hevea rubber and Balata are

TABLE III.

Type of polymer	Ratio A/B	% 1,2 addition
Polybutadiene (emulsion)	.187	21
Polybutadiene (sodium)	1.1	80
GR-S (production)	.180	19

included in this figure for the purpose of comparison.

If the isoprene is polymerized by 1,4 addition, the structure type would be $R_1R_2C=CHR_3$, which corresponds to the structure of Hevea rubber. Compounds of this general type have a characteristic band around 840 cm^{-1} . A band does appear here for the emulsion polymer indicating that this structure is present. In comparing this spectrum with the one for Hevea, it is observed that, although there are some striking similarities, the structures of the two are not identical. However, they do have more in common than do the sodium polymer and Hevea. On this basis, it is reasonable to believe that the emulsion product of polyisoprene is chiefly of the 1,4 structure.

If the polymerization reaction proceeds by 1,2 addition, the resulting compound would fall in the $RCH=CH_2$ class. This is the same general type as the 1,2 addition of butadiene, and strong bands would be expected to appear around 996

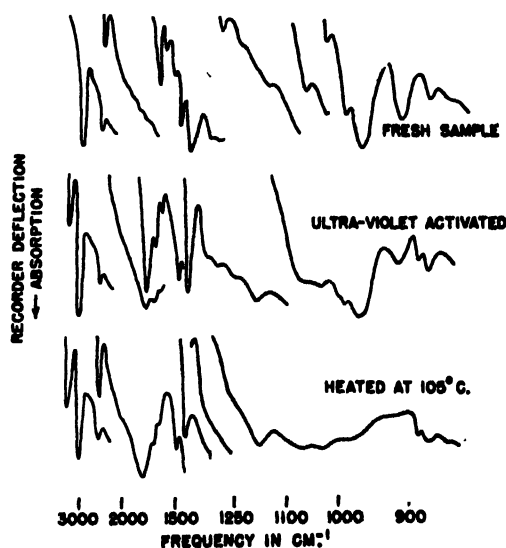


FIG. 6. Absorption curves showing oxidation of GR-S in the absence of anti-oxidant.

cm^{-1} and 914 cm^{-1} . Rather weak bands do appear in these regions for the emulsion polymer while relatively stronger ones are present in the spectrum of the sodium polymer. Thus, a small amount of this structure is indicated in both of these polymers.

The absorption spectrum for the sodium polyisoprene shows a very strong absorption band

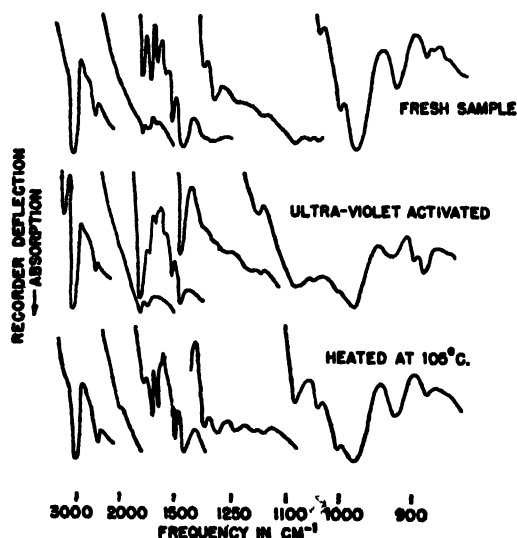


FIG. 7. Absorption curves showing oxidation of GR-S in the presence of anti-oxidant.

at 881 cm^{-1} . Compounds having the general structure $R_1R_2C=CH_2$ are known to have absorption bands in this region. Accordingly, this type of compound would correspond to the 3,4 addition of the polyisoprene structure. The absence of the bands at 840 cm^{-1} and 970 cm^{-1} and the intense absorption at 881 cm^{-1} lead to the belief that the structure of this polymer is predominantly of the 3,4 type.

Balata, the absorption spectrum of which is included in Fig. 5, is generally presumed to be a *trans*-isomer of Hevea rubber. This *cis-trans*-isomerism causes distinctive differences in the infra-red absorption spectra.⁶ Comparison of the spectra for these natural products with those for synthetic polyisoprenes does not disclose the presence of any appreciable amount of the *trans*-isomer in the synthetic polymers. This is particularly interesting, because there has been speculation that the inferiority of the synthetic polymers to Hevea rubber might arise from the fact that a mixture of *cis*- and *trans*-structure occurs in the synthetic polymers.

EFFECTS OF OXIDATION

The effects of exposure of GR-S films to air under conditions such as to accelerate oxidation are shown in the absorption curves appearing in Fig. 6 and Fig. 7. The samples were all prepared

⁶ S. B. Hendricks, S. G. Wildman, and E. J. Jones, Arch. Biochem. 7, No. 3, (Sept. 1945).

from a master solution of production GR-S in benzene from part of which the anti-oxidant (phenyl- β -naphthylamine) was removed by reprecipitation. Films with and without anti-oxidant were exposed to ultraviolet light (Atlas Fadeometer, violet arc) for at least twenty-four hours at a constant temperature of 40°C . Similar films were heated for twenty hours in air at 103°C . The absorption curves resulting for films without antioxidant are shown in Fig. 6. The control or untreated sample is represented in the top curve. The structural changes initiated by exposure to the ultraviolet light are apparent in the middle curve. New bands appear at approximately 3500 cm^{-1} and 1720 cm^{-1} which are definitely due to O-H and C=O linkages, respectively. These groups have been identified chemically in oxidized natural rubber.^{7,8} Increased absorption is also noted in the 1050 cm^{-1} range. The relative intensities of the 996 and 967 cm^{-1} are difficult to interpret. The absorption in the vicinity of the 996 cm^{-1} band is increased but the significance of this is confused by the generally increased absorption over this whole region. The decrease in the intensity of the 914 cm^{-1} band probably is indicative of a definite decrease in the total number of double bonds, which would be expected with oxidation reactions.

When a similar sample was heated for twenty hours in air at 105°C , a number of differences appear in the absorption record when compared with that for the ultraviolet activated sample. This is represented in the bottom curve of Fig. 6. The intensity of the O-H band has increased

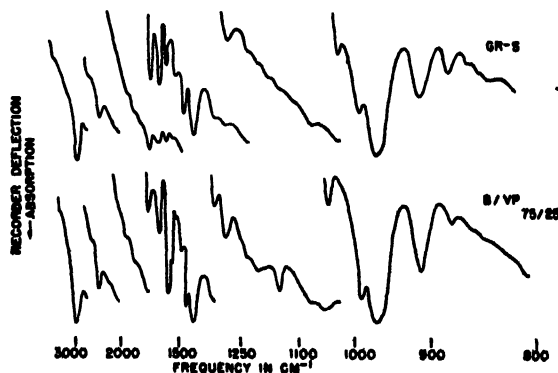


FIG. 8. Absorption curves comparing the structure of GR-S and a copolymer of butadiene and vinyl pyridine.

⁷ R. F. Naylor, Trans. I.R.I. 20, 2, 45 (1944).

⁸ G. F. Bloomfield, Ind. Rubber J. 108, 495 (1945).

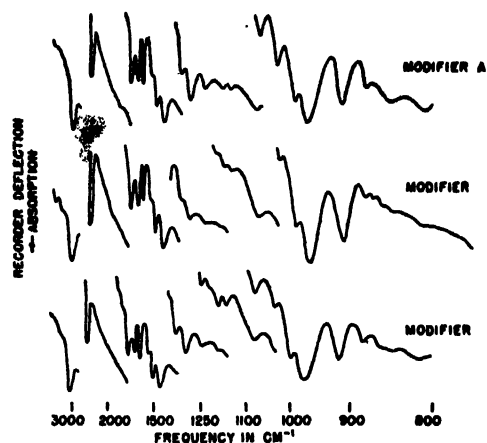


FIG. 9. Absorption curves showing the effect of modifiers on the structure of GR-S.

noticeably. The carbonyl band at 1720 cm^{-1} has broadened considerably. Since the absorptions for the various carbonyls occur at slightly different frequencies, it is possible that the broadening of this band might be due to additional types of carbonyls. The absence of the 914 cm^{-1} band indicates a rather complete saturation of the double bonds. There is a general region of heavy absorption extending from about 900 cm^{-1} to 1300 cm^{-1} , probably indicating heterogeneity in the structure brought about by oxidation. The structure responsible for the occurrence of the 890 cm^{-1} band upon oxidation is not definitely known but its position would suggest the type $R_1R_2C=CH_2$.

In examining the absorption curves of Fig. 7 corresponding to samples with anti-oxidant, it appears that ultraviolet activation is the only agency that has produced any appreciable structural changes. These changes are similar to the ones which occurred for the sample without anti-oxidant. With the anti-oxidant present, twenty hours exposure in air at 105°C produced no noticeable structural changes. It is very interesting that structural changes due to heat treatment are inhibited by the anti-oxidant whereas this anti-oxidant has little influence on the effects caused by ultraviolet activation. Apparently, the mechanisms of oxidation are different in the two cases.

MISCELLANEOUS POLYMER STRUCTURES

An absorption curve for a typical butadiene-vinyl pyridine copolymer is reproduced in Fig. 8

together with one of GR-S for comparison. Although many of the bands are very similar in character, there are several differences that can be observed. One is the relative increase in the intensity of the 1590 cm^{-1} band in the vinyl pyridine copolymer. This band is usually attributed to aromatic double bonded carbons but the presence of the nitrogen in the aromatic group becomes evident with an additional absorption at 1570 cm^{-1} . There is a shift toward the longer wave-lengths of the aromatic CH bending frequencies around 1490 cm^{-1} and a characteristic band at 1150 cm^{-1} . Hence these two synthetic rubbers can be readily distinguished by their infra-red spectra. It is noteworthy that the overlapping bands at 996 cm^{-1} and 967 cm^{-1} are so very similar for the two types of rubber.

Absorption curves are shown in Fig. 9 for samples of GR-S in which different mercaptan

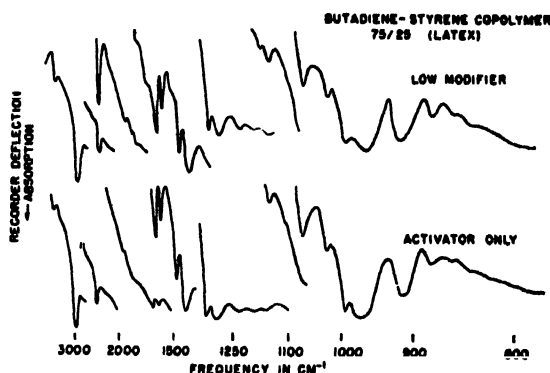


FIG. 10. Absorption curves showing the structure of GR-S in the absence of modifier.

type modifiers were used in the polymerization. These films also contained phenyl- β -naphthylamine as anti-oxidant. Since the presence of modifiers during polymerization influences the physical properties of the product, it was thought that structural differences might be detected in these absorption spectra. However, the absorption curves for polymers made with these three modifiers are strikingly similar. Since the 996 cm^{-1} band shows no evident difference in the number of side vinyl groups present and other differences in the spectra are insignificant, it can be concluded that these three modifiers influence the structure in much the same way, at least as far as can be detected by infra-red methods. In this case, there appears to be little hope that

these methods will detect such differences, since the fundamental structure is so similar. The differences in physical properties evidently result either from a variation in a relatively small proportion of the chemical linkages, as might readily be the case for cross linking, or in variations which do not give rise to unique absorption bands. The amount of modifier present is so small that it does not appear directly in any of the absorption curves.

Since unmodified copolymers of butadiene and styrene are generally insoluble in organic solvents, films were made from latices of the copolymers containing little or no modifier. The absorption curves for these are shown in Fig. 10, respectively for samples containing .1 percent mercaptan activator-modifier and .1 percent mercaptan activator. The latter activator is considered to have little or no modifying action. The increased

number of side vinyl groups as compared to GR-S becomes apparent from the increased absorption of the 996 cm^{-1} band. This increase is more pronounced for the slightly modified sample than for the one which contained activator only. This evidence is far from conclusive as the effect of the modifier and activator on the amount of 1,2 polymerization but indications are that when the normal amount of modifier is used there are less side vinyl groups formed than if no modifier or only a very slight amount of modifier is present.

We wish to express our appreciation to Dr. L. B. Sebrell and The Goodyear Tire & Rubber Company for permission to publish this work, to Dr. A. M. Borders and Dr. J. O. Cole of the Research Staff for preparing the polymers used, and to Dr. M. R. Fenske for supplying the samples of octene-1 and octene-2.

A Photoelectric Instrument for Light Scattering Measurements and a Differential Refractometer

P. P. DEBYE*

General Aniline and Film Corporation, Central Research Laboratory, Easton, Pennsylvania

The method of determining particle sizes and molecular weights by light scattering measurements on solutions is finding increasing fields of application. In order to obtain the necessary data quickly and conveniently, an instrument has been developed which allows the required light scattering measurements to be taken on a routine basis. This apparatus is described. It measures transmission and the scattering of light at 90° ; two different schemes can be employed to obtain the angular intensity distribution of the scattered light. A description is also given of a compensating refractometer which is used to determine the difference in refractive index of solvent and solution. The

instrument, which is simple in design, gives a direct reading of the refractive index difference and has an accuracy of better than 10^{-5} . The light scattering instrument has been tested using polystyrene of known molecular weight, and has been applied to study the change of apparent molecular weight (or particle size) of a series of polyvinyl ethers as a function of concentration in some cases, also a number of solvents. For a series of polyvinyl *n*-butyl ether samples which was studied, the relation between the intrinsic viscosity $[\eta]$ of the benzene solutions and the weight average molecular weight M determined by light scattering is $[\eta] = .365 \times 10^{-4} M^{1/2}$.

In 1944, P. Debye¹ described a method for determining by light scattering measurements on solutions the molecular weight and particle size of the solute. Since then this procedure has become an important tool in the field of high polymers, and it seemed desirable to have special instruments to make the necessary measurements more conveniently.

The fundamental concepts of the theory show that the diameter of particles larger than about $1/20$ of, but still comparable to, the wave-length of the light can be determined if the angular intensity distribution of the scattered light is known. To illustrate this Fig. 1 shows the intensity of the scattered light as a function of the angle; the family of curves is drawn for varying ratios d/λ , where d is the particle diameter and λ the wave-length of the light in the solution. This particular angular intensity dis-

* Now at Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.
¹ P. Debye, J. App. Phys. 15, 338 (1944).

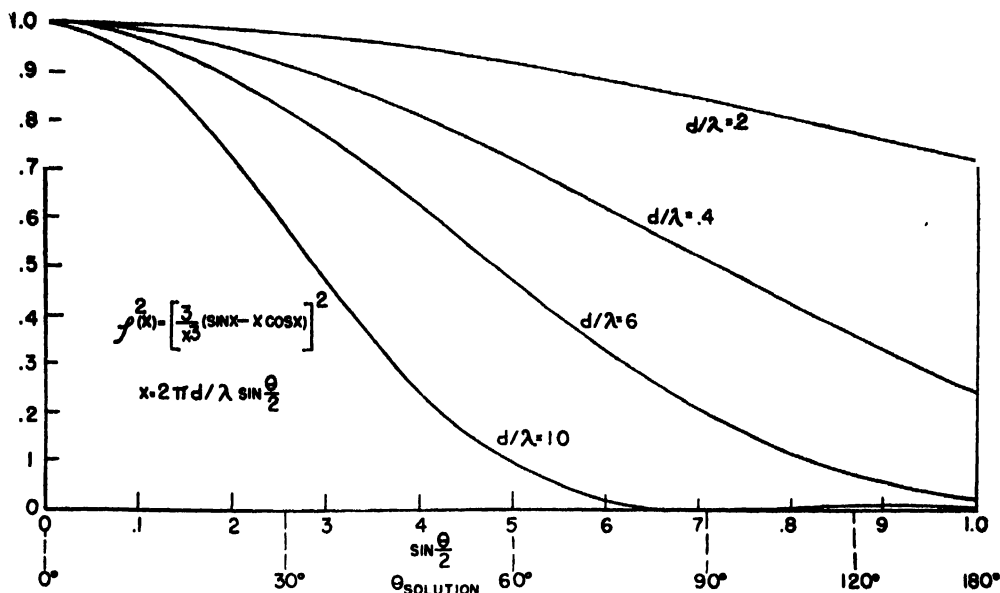


FIG. 1. Theoretical angular distributions for spherical particles.

tribution has been calculated for spherical particles and the angular intensity variation due to the polarization effect has been omitted.

In case the particles are smaller than approximately $1/20$ of the wave-length and the solution scatters as much in the forward as in the backward direction, P. Debye's formula for the turbidity (τ) of the solution (in the limit for infinite dilution) is applicable:

$$\tau = \frac{32\pi^3}{3} \frac{\mu_0^2(\mu - \mu_0)^2}{\lambda^4 n},$$

where μ_0 and μ are the refractive indices of solvent and solution, respectively, λ is the wave-length of the light in vacuum and n the number of particles per cc. For practical purposes Avogadro's number (N), the molecular weight (M), and concentration (c) in g/cc can be introduced leading to the relation

$$Kc/\tau = 1/M,$$

$$K = \frac{32\pi^3}{3} \frac{\mu_0^2}{N\lambda^4} \left(\frac{\mu - \mu_0}{c} \right)^2.$$

From this the number (n) of solute particles per cc can be calculated. The turbidity in practical concentrations and solvents is for many

materials quite small. In most cases it is therefore easier to measure the amount of scattered light, for example as 90° scattering, instead of the small difference in intensity of the primary light beam before and after it passed through the solution. The quantity of $\mu - \mu_0$ also requires careful measurement; its magnitude is in general approximately .001 for a 1 percent solution. The useful lower molecular weight limit of the method is determined by the refractive index difference that can be obtained for a given polymer-solvent system. It lies roughly in the neighborhood of a molecular weight of 10,000. The weights as determined are weight average molecular weights.

The two instruments² presently to be described were developed in order to measure (a) the angular intensity distribution, (b) turbidity by the measurement of transmission, (c) turbidity by measurement of 90° scattering, and (d) the refractive index difference ($\mu - \mu_0$) of solution and solvent. The photoelectric instrument is adapted for measurement of a , b , and c and the differential refractometer reads $\mu - \mu_0$ directly.

The light scattering unit with its amplifier is shown in Fig. 2. The unit contains four separate

² Both instruments were built by Mr. H. Bush, Chemistry Department, Cornell University, Ithaca, New York.



FIG. 2.

photo-tubes of the type RCA 929 which can be connected interchangeably to a special d.c. amplifier.³ The light intensities are read on a built-in taut suspension galvanometer, which is shunted in order to have a variable sensitivity range of 1:2:5:10:20:50. The linearity of this arrangement has been checked on a photometer bench and has been proved satisfactory. Before and after each series of measurements the primary light beam intensity is obtained by reflection on a magnesium carbonate block under 45 degrees with respect to the primary light beam and with an additional dark glass filter interposed. In this way long term fluctuations of the light source as well as of the amplifier can be eliminated.

The light scattering unit has three separate and easily interchangeable lamp houses; one house containing a high intensity tungsten filament bulb, the other a water-cooled high pressure mercury arc (H-6), and the third a medium pressure-forced air cooled mercury arc of the AH4 type. Almost all measurements are taken with this latter lamp. It has been found that a simple electrical stabilizing circuit makes the light output of the AH4 bulb quite satisfactorily constant. The circuit used is shown in Fig. 3; it consists of a voltage regulating autotransformer (T_1), a current regulating ballast lamp (B), a step-up transformer (T_2) to compensate for the voltage drop in the ballast, and the auxiliary transformer (T_3) which is always necessary to operate the AH4 bulb from line current.

The arrangement of the optical parts is shown schematically in Fig. 4. The light of the mercury

arc illuminates an opal glass (O) situated in front of a slide (A), having various interchangeable apertures of different sizes; then it passes a monochromatic glass color filter (F) and a stationary lens (L_2) with iris diaphragm (D). The parallel beam of light which leaves the lens can then be intercepted by a Polaroid (P) and/or an extra plain glass slide (G). The latter is used to calibrate the electrical parts by decreasing the primary light intensity by small but known amounts. This is useful in the case of transmission measurements where the zero of the galvanometer is suppressed by electrical means. Between the aperture slide and the color filter there is a removable additional lens (L_1) which on account of its higher numerical aperture gives greater light intensities. It focuses the first aperture in the center of the solution cell (S).

As shown in Fig. 4 one photo-tube is situated close to the solution cell for measuring turbidity by 90° scattering. The other photo-tube is used for transmission measurements but it can also be applied to measure angular distribution; the latter is done by moving the tube in a semi-circle around the solution cell. The perpendicular axis of this motion extends through the bottom of the unit and is connected to an indicator arm on the outside of the case, where a pointer shows the position of the tube in degrees with respect to the primary beam. The motion covers an angle interval from 0 to 125 degrees.

In an arrangement like this it is necessary to employ a circular solution cell unless one applies corrections for the varying reflection and absorption, which is at least time consuming. In order to avoid this a suggestion by P. Debye has been followed. The same support which carries the photo-tube for transmission measurements carries in addition another photo-tube which moves on a semi-circle underneath the solution cell. This tube receives its light from a mirror (at 45° to the horizontal) which travels suspended from

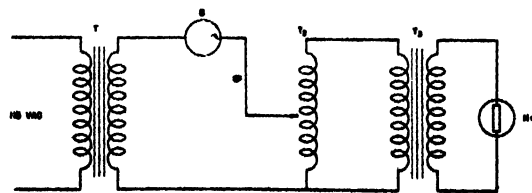


FIG. 3. Light source stabilization circuit.

³ Photovolt Corporation, 95 Madison Avenue, New York, New York.

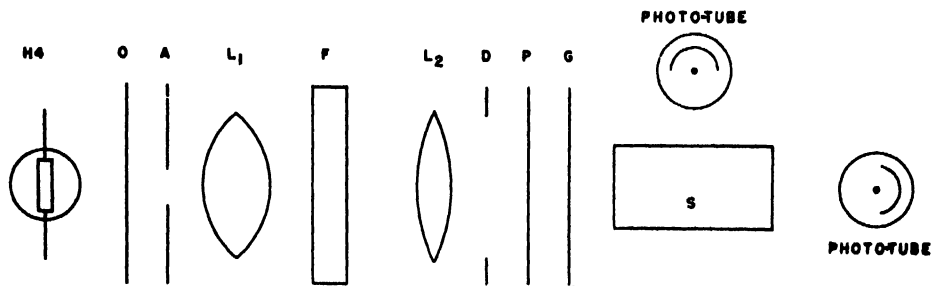


FIG. 4. Schematic arrangement of optical parts in the light scattering unit.

the outside but within the solution on the same semi-circle as the photo-tube. A sketch of this arrangement is shown in Fig. 5. By this method the scattered light always leaves the solution cell perpendicularly to the interface solution-air and the corrections otherwise necessary are avoided

A fourth photo-tube not shown in the diagram is used to measure independently the intensity of the light source, it receives the light from a plain glass which is placed under 45 degrees in the primary beam

The apparatus has been calibrated with a solution of polystyrene of which the weight average molecular weight had been determined as $409,000 \pm 7$ percent. This figure has been arrived at by J. McCartney⁴ who used an instrument of special design to measure the turbidity. The reported number average molecular weight by osmometry for this polystyrene was 240,000.

The differential refractometer (Fig. 6) allows the measurement of $\mu - \mu_0$ with great convenience and an accuracy of about 0.000,003. This is

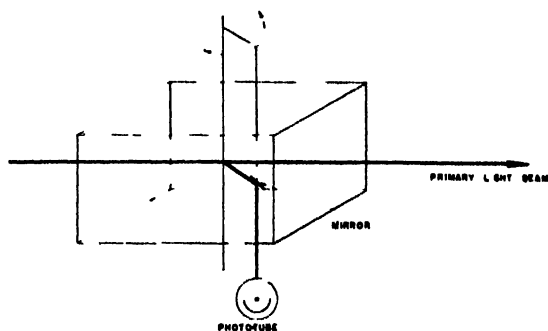


FIG. 5. Schematic arrangement for the measurement of the angular intensity distribution.

⁴J. McCartney, Chemistry Department, Cornell University, Ithaca, New York.

appreciably better than can be achieved with the ordinary type of refractometer and not quite as good as the accuracy possible with an interferometer. However, the great simplicity of the instrument and the ease with which it performs makes this refractometer a great asset if a number of refractive index differences have to be obtained every day. The original design was developed in the course of work for the Rubber Reserve Corporation at the Chemistry Department, Cornell University, in 1943; the instrument as described here is an improved copy of the Cornell refractometer.

A schematic drawing is shown in Fig. 7. An adjustable precision slit (S) is illuminated by monochromatic light from a mercury arc (AH-4) and an image (I) of this slit is formed at a distance of about 190 cm in the field of a filar micrometer. A rectangular glass cell (C) containing the solvent is placed between the two lenses (L_1 and L_2) in the parallel light beam. Suspended in the solvent cell (C) is a hollow prism glass cell (P) containing the solution. The refraction angle of the prism is approximately 125 degrees. If there is any difference of refractive index of solvent and solution, the image of the slit which can be observed in the eyepiece of the filar micrometer will undergo a deflection. If the

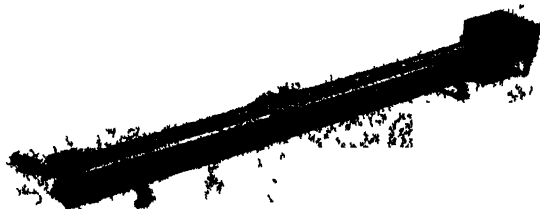


FIG. 6.

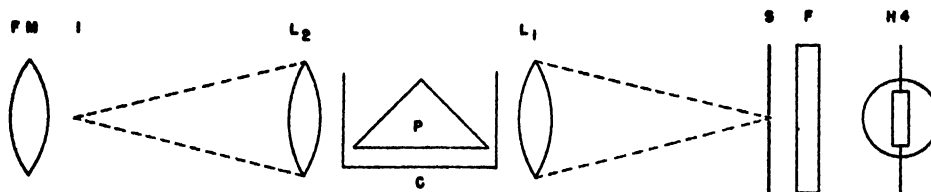


FIG. 7. Schematic arrangement of the differential refractometer.

difference in refractive index ($\mu - \mu_0$) is not too large, the image deflection will be proportional to the refractive index difference. For this instrument the full scale deflection is still in the linear range.

The image deflection can be calculated by the following equations. If the top angle of the prism is A , the angular deflection α and prism and cell have the refractive indices μ and μ_0 , respectively, then

$$\alpha = 2(\mu - \mu_0) \tan A/2.$$

With a lens of focal length (f) this gives an image deflection X

$$X = \alpha f = 2f \tan A/2 (\mu - \mu_0).$$

The relation shows that the image deflection X depends linearly on the refraction difference $\mu - \mu_0$ only in first approximation (for small deflections).

The equation for the image deflection (X) also shows that X is independent of the refractive index of the solvent (μ_0). The instrument can therefore be calibrated with, for instance, an aqueous sucrose solution and then be used for any other solvent-solute system.

The differential method employed in this instrument holds the great advantage of being

fairly insensitive to small temperature changes. On this account a good enclosed type mercury regulator will work well for the temperature control of the water which is pumped through the jacket surrounding the cells (Fig. 8). In practice the temperature equilibrium of prism and cell is attained in 5-7 minutes.

The refractometer has been calibrated with dilute sucrose solutions of known refractive index. The calibration (Fig. 9) is shown to be linear.

Since the completion of these instruments, a great number of materials have been tested. Some of the experimental results on polystyrenes are shown in Fig. 10. The data for Kc/τ have been linearly extrapolated (by the method of least squares) to zero concentration, at which point the ordinate corresponds to the reciprocal of the actual molecular weight. The scale in Figs. 10-12 is drawn in arbitrary units.

Another group of polymers which has been of great interest has been that of the polyvinyl ethers. Figure 11 shows the experimental curves for various polyvinyl normal butyl ethers. It is interesting to note that the plotted data for these materials show a definite downward curvature.



FIG. 8.

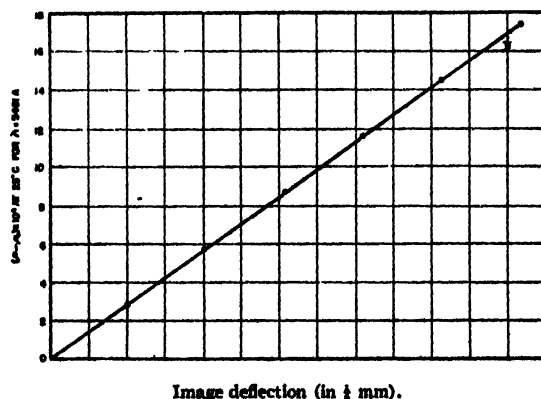


FIG. 9. Calibration of differential refractometer.

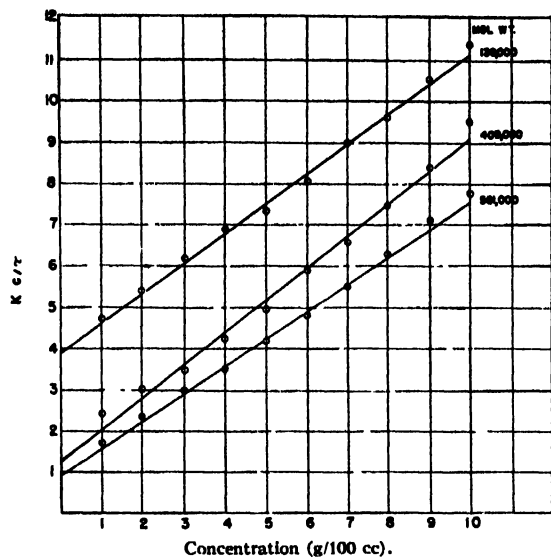


FIG. 10. Experimental data on polystyrenes in toluene.

This might be explained by agglomeration of particles at increased concentration. A very broad particle size distribution also would show a similar effect. In this case the purpose of the investigation was to connect molecular weights with viscosity values in order to obtain the constants in the familiar expression

$$\log \eta_s = A + \alpha \log M.$$

The intrinsic viscosities were determined in benzene solution at 25 C. The flow times were measured in tubes of the modified Ostwald-Fenske design. The constants for the polyvinyl normal butyl ether were determined as $A = -7.14$ and $\alpha = 1.36$.

Another interesting case was found in a preliminary study of the effect of milling (at 25–30 C) on a sample of polyvinyl isobutyl ether. Figure 12 shows that the slopes of the Kc/τ vs. concentration data were very much decreased by milling the polymer. Considering the experimental error involved, however, this procedure did not seem to affect appreciably the weight

TABLE I.

Material	$\eta_{sp}/\text{concentration}$ in benzene at 25 C for conc. = .2 g/100 cc	Molecular weight by light scattering
Not milled	1.75	219,000
10 min. milled	1.01	233,000
45 min. milled	.68	240,000

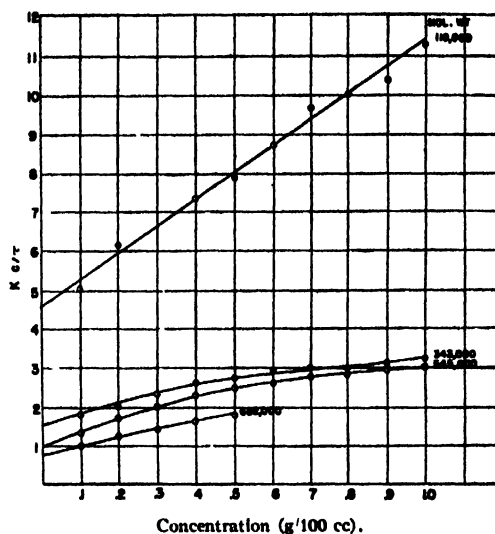


FIG. 11. Experimental data on polyvinyl *n*-butyl ethers in benzene.

average molecular weight of the materials under the conditions of this experiment. This might indicate that the homogeneity of the sample increases by milling, which necessitates not only a breaking down of the high molecular weight fraction but also at the same time a building up of larger units from the lower molecular weight fraction. The smaller slope of the milled samples could also be explained by a decreased solubility. The milled polymer may have been oxidized during the milling process resulting in a less soluble material. At present this question is being studied with polymers that have been milled in inert atmosphere.

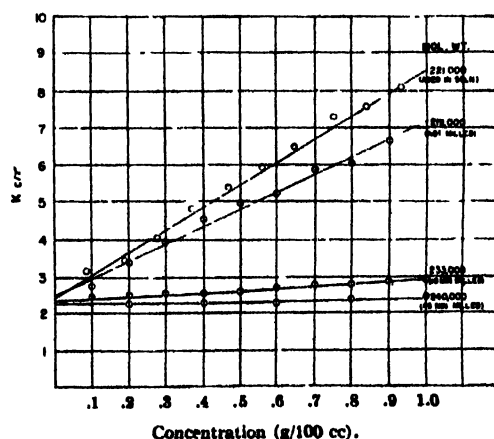


FIG. 12. Experimental data on polyvinylisobutylether in benzene.

The relatively small change in the average molecular weight is surprising if one considers the decrease in specific viscosity/concentration. A comparison of viscosity data and the molecular weight as determined by light scattering is shown in Table I.

Figure 12 also shows as another example the

experimental data for the unmilled material after it has been "aged" in solution for 48 hr. at about 65 C. Again the molecular weight is essentially unchanged. However, the greater slope seems to indicate a wider particle size distribution or a change in solubility. Further work should be done to extend these experiments.

Thermal Expansion and Second-Order Transition Effects in High Polymers

III. Time Effects

R. S. SPENCER AND R. F. BOYER

Physical Research Laboratory, The Dow Chemical Company, Midland, Michigan

It has been suggested that the so-called second-order transition in high polymers is not an equilibrium phenomenon, i.e., a true thermodynamic singularity, but rather a rate effect. Confirmation of this point of view has been obtained by determining the equilibrium volume-temperature curve of polystyrene. Such equilibrium curves exhibit no transition within the range of from 20° to 140°C, whereas the same material shows a transition at about 82°C for rates of heating usual in thermal expansion measurements. Two facts appear in agreement with the viscous flow interpretation of the thermal expansion process: First, two mechanisms were found to operate, at markedly different rates. One gave almost instantaneous expansion, even at room temperature, whereas the other was quite temperature-dependent in this region, being extremely slow at room temperature. Second, the activation energy for the slower mechanism was found to be of the order of magnitude of that for viscous flow, considerably lower than the activation energy for rubber-like elasticity. Volume-temperature curves at finite rates of heating are discussed in light of these findings.

INTRODUCTION

THE thermal expansion of high polymers exhibits features quite different from the usual behavior of simple organic liquids. The volume-temperature curve is characterized by the presence of a temperature, or interval of temperature, which separates the curve into two portions corresponding to two markedly different expansion coefficients. This temperature has been known as the second-order transition temperature, so-called because of the fact that there is an apparent discontinuity in the secondary thermodynamical quantities, such as expansion coefficient, heat capacity, etc., at that point.² As the temperature is raised through this point, there appears to be a change of state, without

latent heat, resulting in a larger value of thermal expansion coefficient and heat capacity and a change in the general physical properties of the material. The polymer softens and acquires rubber-like elastic properties. Significantly enough, however, there is no discontinuity of first- or second-order in the viscosity.³

Several interpretations of this phenomenon have been suggested, with no very conclusive evidence appearing, as yet, to confirm any one of them completely. The second-order transition might be a true thermodynamical singularity of the type discussed by Ehrenfest⁴ and Mayer and Streeter.⁵ Justi and Laue⁶ have pointed out that Ehrenfest's conditions for a second-order transition are not sufficient, and the further condition

¹ R. F. Boyer and R. S. Spencer, *J. App. Phys.* 15, 398 (1944).

² R. F. Boyer and R. S. Spencer, *J. App. Phys.* 16, 594 (1945). The first two references present a brief survey of second-order transition phenomena and give a number of references to earlier work.

³ G. S. Parks, *et al.*, *Physics* 5, 193 (1934).

⁴ P. S. Ehrenfest, *Comm. Leiden, Suppl.* 75b (1933).

⁵ J. E. Mayer and S. F. Streeter, *J. Chem. Phys.* 7, 1019 (1939).

⁶ E. Justi and M. Laue, *Zeits. f. tech. Physik* 12, 521 (1934).

they proposed would lead to something which should perhaps more properly be termed a third-order transition. Pines⁷ has extended the thermodynamical treatment to the case in which the transition appears to take place over a temperature interval rather than at a single point. The authors have suggested elsewhere⁸ the possibility that the apparent transition in high polymers might consist of a series of λ -transitions, of the type encountered in some simpler molecules,⁵ occurring at temperatures spaced quite close together. In this way the general features of the heat capacity-temperature curve may be accounted for.

The other viewpoint, which is now rather widely accepted, is to assume that there is no thermodynamical singularity at the second-order transition and to ascribe all attending phenomena to the temperature dependence of the time-rate of whatever internal mechanism is operating to give the macroscopic property under observation. Thus Alfrey, Goldfinger, and Mark⁹ interpret the apparent second-order transition as the point where the rate of attainment of equilibrium is of the same order of magnitude as the time-scale of ordinary physical experiments. In the case of the change of volume with temperature, they distinguish between two mechanisms of expansion: (1) an over-all increase in intermolecular distances due to enhanced thermal vibrations, as in the case of a crystalline solid, and (2) the diffusion of "holes" through the material.

The rate of the second process is strongly temperature-dependent and is taken to account for the transition phenomena observed. The authors have proposed² a similar scheme in which more emphasis is placed on the viscous flow of chain segments necessary to create new "holes" and thereby increase the volume. The second-order transition then begins at the lowest temperature at which an observable volume change is produced by viscous flow of the segments under the expansive forces acting on the segments, within the time limits of the experi-

mental technique employed. This introduction of the concept of viscous flow into the thermal expansion process has proven very fruitful in interpreting and correlating the effects of various external variables, such as applied stresses, plasticizers, and experimental time scale, on the transition temperature. It has also served to explain the origin of the second-order transition observed in heat capacity measurements, which will be discussed in more detail in a later paper in this series.

Available data seem to favor regarding apparent second-order transitions as rate effects rather than thermodynamical singularities. Jenckel¹⁰ was able to lower the transition temperature of selenium glass 7°C by extending the heating time from several hours to several days. Alfrey, Goldfinger, and Mark⁹ also reported a 7°C variation of the transition temperature of polystyrene as the rate of heating or cooling was varied. Richards¹¹ has pointed out that the transition temperature of glucose glasses varies over a much wider interval, from 7° to 77°C, as the rate of observation is speeded up in the sequence heat capacity-thermal expansion-1000 c.p.s. dielectric constant.

On the other side of the picture it should be pointed out that no one has, as yet, reported data extending the volume-temperature curve found above the transition temperature (supposedly representing a state of equilibrium) below the transition temperature observed at usual heating rates. Winter's work on glass¹² suggests that a similar anomaly in refractive index is not entirely a rate effect, but rather that the transition persists even under conditions of equilibrium. She selected several temperatures in the transition region and at each temperature determined the isothermal refractive index-time curve with the glass initially in the state above the transition region and also with the glass initially in the state below the transition region. The limiting refractive index defined by these two curves was found to be identical and was taken as the equilibrium refractive index at the specified temperature. In this way the equilibrium refractive index-temperature curve was

⁷ B. J. Pines, *J. Exp. and Theo. Phys. (U.S.S.R.)* 9, 963 (1939).

⁸ R. F. Boyer and R. S. Spencer, *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. II.

⁹ T. Alfrey, G. Goldfinger, and H. Mark, *J. App. Phys.* 14, 700 (1943).

¹⁰ E. Jenckel, *Zeits. f. Elektrochemie* 43, 796 (1937).

¹¹ W. T. Richards, *J. Chem. Phys.* 4, 449 (1936).

¹² A. Winter, *J. Am. Ceram. Soc.* 26, 189 (1943).

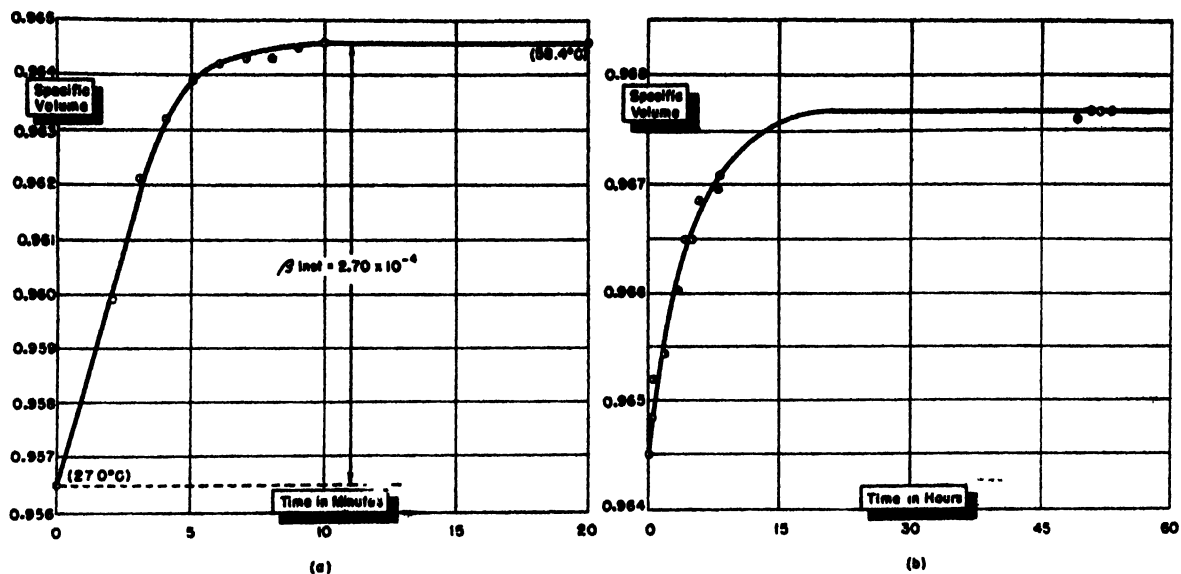


FIG. 1. Isothermal expansion of polystyrene at 58.4°C. (a) Thermal lag in dilatometer and "instantaneous" expansion. (b) Slow approach to equilibrium.

obtained, and it was found that the transition did not disappear but remained, even at equilibrium. For glass, therefore, the second-order transition seems to be an essential singularity

with superposed rate effects. This conclusion prompted the following work on polystyrene, in which the equilibrium volume-temperature curve was determined.

EQUILIBRIUM VOLUME-TEMPERATURE CURVE OF POLYSTYRENE

The polystyrene used in this study was in pellet form, had a molecular weight of 300,000, and had been stored at room temperature for approximately two years. This made it seem reasonable to assume that this material was very close to its equilibrium volume at room temperature. The technique of handling samples was similar to that described in an earlier paper.¹ A weighed sample of polymer was placed in the bulb of a Pyrex dilatometer, followed by a hollow glass plug, and the bulb was sealed off. The dilatometer was then evacuated and filled with a known amount of mercury through the capillary tube. The position of the mercury in the capillary tube at room temperature was noted.

The dilatometer was next transferred to a constant temperature glycol bath and the height of the mercury in the capillary tube recorded as a function of time, over a period of several days. When it was evident that equilibrium had been reached, the dilatometer was transferred to a bath held at a higher temperature and the process

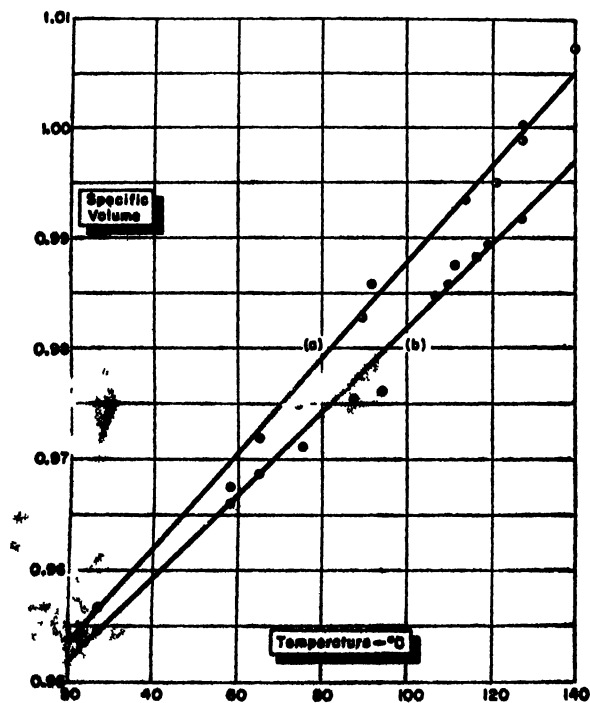


FIG. 2. Equilibrium volume-temperature curves of two samples of polystyrene. (a) $\beta_{\text{inst.}} = 2.70 \times 10^{-4}$ per °C, $\beta_{\text{eq.}} = 4.53 \times 10^{-4}$ per °C. (b) $\beta_{\text{inst.}} = 2.83 \times 10^{-4}$ per °C, $\beta_{\text{eq.}} = 3.98 \times 10^{-4}$ per °C.

repeated. Occasionally, after the highest desired temperature had been reached, some of the points were re-checked on cooling.

It was found that when a polymer sample was suddenly raised in temperature, it expanded to a new volume almost instantaneously, being slowed down somewhat, of course, by thermal lag; after this initial expansion the material continued to expand, isothermally, at a slow rate until equilibrium was reached, within the limits of observation. If the sample were cooled, instead of heated, the same thing happened, except that the material contracted instead of expanding. The length of time necessary to reach equilibrium was dependent on the temperature, higher temperature resulting in shorter time.

This behavior is shown in Fig. 1, which is typical of the results obtained. In 1(a) we see the rapid change in volume associated with the thermal lag in the dilatometer bulb; and in 1(b) we see the much slower drift of volume toward equilibrium, following the attainment of thermal equilibrium. The pseudo-equilibrium attained in 1(a) corresponds to the "instantaneous" expansion of the sample, whereas that of 1(b) is a true equilibrium volume.

The equilibrium volumes of a polystyrene sample were determined at different temperatures and plotted as the equilibrium volume-temperature curve. Figure 2 shows such equilibrium volume-temperature curves for two different samples of polystyrene. It may be seen that from room temperature upwards the volume is essentially a linear function of the temperature, there being no discontinuities or transitions. The instantaneous (β_s) and equilibrium (β_e) cubical expansion coefficients (referred to the volume at 0°C) were slightly different for the two samples. They correspond roughly in magnitude to the expansion coefficients reported by Wiley¹³

TABLE I. Half-times for isothermal expansion and contraction of polystyrene.

Temperature	Half-time
30.0°C	13.6 hours
58.7	2.9
65.4	1.32
74.5	0.94
90.0	0.41

¹³ F. E. Wiley, Ind. Eng. Chem. **34**, 1052 (1942).

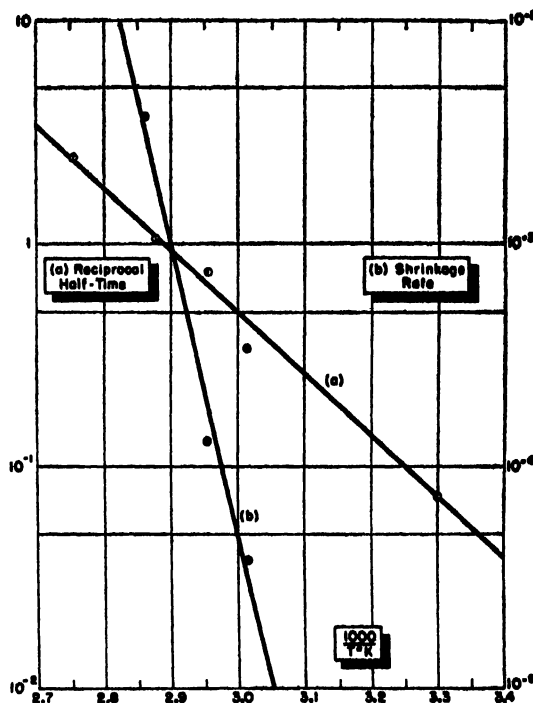


FIG. 3. Effect of temperature on the rates of two molecular processes in polystyrene. (a) Half-times of isothermal volume changes. Activation energy = 12 kcal. (b) Shrinkage rate of an oriented specimen. Activation energy = 61 kcal.

and by Alfrey, Goldfinger, and Mark⁹ for polystyrene below and above its transition temperature, respectively. The points represented include those obtained by both isothermal expansion and contraction. No hysteresis was observed.

An unsuccessful attempt was made to carry out the same type of study on polymethyl methacrylate and a copolymer of styrene and divinyl benzene. The polymethyl methacrylate appeared to release gas upon heating, which rendered it impossible to use the dilatometer method. The styrene-divinyl benzene copolymer seemed to have residual strains at room temperature, which gradually released upon heating, resulting in anomalous volume-time curves.

NON-EQUILIBRIUM PHENOMENA

The authors have suggested² the relationship

$$\tanh [\alpha(V - V_e)/2] = Ce^{-kt} \quad (1)$$

for isothermal volume changes in high polymers, where V is the volume at time t , V_e is the equilibrium volume, and α , C , and k are constants.

The volume-time curves obtained in this study were found to conform fairly well to the above relationship, but it was also noted that they fitted equally well a simple exponential relationship, as would be expected for relatively small values (≤ 1) of the argument of the hyperbolic tangent. This observation suggested the possibility of determining the activation energy of the process of retarded thermal expansion. Table I lists half-times of the volume-time curves for several temperatures, and Fig. 3 shows these data plotted in the usual semi-logarithmic manner (curve (a)). It is seen that a fairly straight line results, corresponding to an activation energy of 12.0 kcal. for the process. It is interesting to compare this value with the activation energy for viscous flow in polystyrene and that for high or rubber-like elasticity. It is in best agreement with an activation energy for viscous flow of polystyrene of 12 kcal., reported by Ferry¹⁴ and obtained by studying the viscosities of solutions in xylene and extrapolating to 100 percent polystyrene. A similar study (mentioned in reference 2) of the viscosities of solutions of polystyrene in isopropyl benzene gave an activation energy of 9.9 kcal. by extrapolation to 100 percent polymer. In contrast to this, Foote¹⁵ reports a value of 40 kcal. and the data of Wiley¹⁶ gives an activation energy of 31.6 kcal. However, Tuckett¹⁷ has pointed out that viscous and elastic effects have not been separated in measurements such as those of Wiley, resulting in an activation energy higher than that for pure viscous flow. He further suggests that the activation energy for viscous flow in linear polymers without special bonds should be of the order of 10–30 kcal. and that higher values, 50–100 kcal. are usually due to elastic effects.

In this connection it was thought to be of interest to determine the activation energy of rubber-like elasticity in polystyrene. A rod of polystyrene was heated, stretched to about four times its original length, and a specimen cut from its center. This specimen was placed in a quartz dilatometer of the type recommended in

the ASTM test D696-42T and the initial rate of shrinkage parallel to the orientation axis determined at three temperatures. The actual shrinkage observed was so small, as compared to the total stretch, that the length-time curves were essentially linear, and no correction was made for the slightly different initial length at each temperature. Curve (b) of Fig. 3 shows a semi-logarithmic plot of rate of shrinkage against reciprocal absolute temperature, from which an activation energy of about 61 kcal. is estimated. It is quite apparent from these values that the thermal expansion process in polystyrene does not involve any very great amount of chain uncoiling, but rather viscous flow of chain segments.

Thermal expansion measurements are not usually carried out under equilibrium conditions, and it is of interest to consider the volume-temperature curves resulting from various heating conditions and initial states. Under non-equilibrium conditions, i.e., at finite rates of temperature variation, the volume changes by two mechanisms

$$d(\Delta\sigma)/dt = -B\Delta\sigma, \quad (2)$$

$$d(\Delta\sigma)/dt = -\sigma_e(273)(\Delta\beta)dT/dt, \quad (3)$$

where $\Delta\sigma$ is the difference between the actual specific volume at time t and temperature T and the equilibrium specific volume at T , B is a quantity of the form $A \exp(-E/RT)$ which may be obtained from Fig. 3, $\sigma_e(273)$ is the equilibrium specific volume at 0°C, and $\Delta\beta$ is the difference between the equilibrium and "instantaneous" cubical expansion coefficients. Combining these two equations and making temperature the independent variable, we have

$$\frac{d(\Delta\sigma)}{dt} = -\frac{B\Delta\sigma}{dT/dt} - \sigma_e(273)\Delta\beta, \quad (4)$$

where dT/dt is now to be considered as a function of T . For given initial conditions and a given temperature-time curve a graphical solution of this equation is readily obtainable by the method of isoclines.¹⁸ This method is quite flexible, for

¹⁴ J. D. Ferry, J. Am. Chem. Soc. 64, 1330 (1942).

¹⁵ N. M. Foote, Ind. Eng. Chem. 36, 244 (1944).

¹⁶ F. E. Wiley, Ind. Eng. Chem. 33, 1377 (1941).

¹⁷ R. F. Tuckett, Trans. Faraday Soc. 39, 158 (1943).

¹⁸ T. K. Sherwood and C. E. Reed, *Applied Mathematics in Chemical Engineering* (McGraw-Hill Book Company, Inc., New York, 1939), p.126. W. C. Johnson, *Mathematical and Physical Principles of Engineering Analysis* (McGraw-Hill Book Company, Inc. New York, 1944), p. 86.

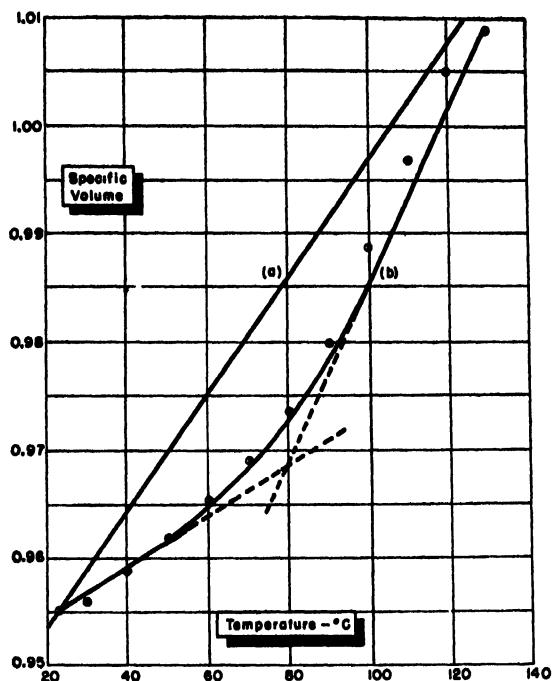


FIG. 4. Expansion of polystyrene from equilibrium at 23°C, with a heating rate of 0.82°C per minute. The circled dots are the experimental points. (a) Equilibrium curve for comparison. (b) Theoretical curve under the above conditions.

one family of curves ($\Delta\sigma$ versus T), corresponding to constant values of the product $B\Delta\sigma$, will suffice for any initial conditions and any temperature schedule. A comparison between theory and experiment is made in Fig. 4, where the equilibrium curve is indicated by the curve (a), the theoretical curve for constant rate of temperature rise by curve (b), and the observed curve under the same conditions by the circled points. The general features of the theoretical and experimental curves were the same, and a reasonably satisfactory fit was obtained. It is interesting to note that the system never reaches equilibrium, but rather the actual curve approaches the equilibrium curve asymptotically at the higher temperatures. Also, the non-equilibrium expansion coefficient above the apparent transition is actually greater than the equilibrium value.

Another case of interest is the thermal expansion of a quenched sample. By this we mean that a sample at equilibrium at some elevated temperature is cooled at such a rapid rate as to permit contraction corresponding to only the

instantaneous expansion coefficient. The material is then immediately heated at a constant rate and the volume-temperature curve observed. The theory predicts that the volume will start out above the equilibrium curve and eventually cross it at a temperature a little below the temperature from which the sample was quenched, this temperature difference depending on the rate of heating. The volume then will dip below the equilibrium curve, to an extent depending on the heating rate, and approach the equilibrium curve asymptotically from below.

The importance of these conclusions in practical considerations such as the question of the dimensional stability of molded polystyrene objects is at once apparent. If the molded object is at equilibrium volume at some temperature and is then removed from the mold and cooled rapidly, the volume upon reaching room temperature is somewhat greater than the equilibrium volume. The object will then shrink over a period of several days, gradually approaching the equilibrium volume. The amount of shrinkage is roughly proportional to the temperature interval covered during the rapid cooling. The expansion coefficients of Fig. 2, for example, would result in a volume contraction of about 1.7 percent for a 100°C cooling interval. This shrinkage at room temperature may be reduced by an annealing process, i.e., by one or both of two procedures: first, lowering the annealing temperature, that is to say, the temperature at which the volume is allowed to come to equilibrium, and second, going to slower cooling. Graphical solution of Eq. (3) for various annealing conditions, as before indicated, permits the determination of the optimum annealing schedule for a given, allowable amount of shrinkage.

As an example of the type of chart which might prove useful in annealing problems, Fig. 5 shows the effect of temperature on the time interval necessary for the volume of a polystyrene sample to cover 50, 75, 90, 99, and 99.9 percent of its distance from the equilibrium volume, by isothermal expansion or contraction. Such a family of curves would facilitate the choice of an annealing time to correspond to any given annealing temperature. Similarly, curves might be constructed showing the deviation from equilibrium at room temperature after

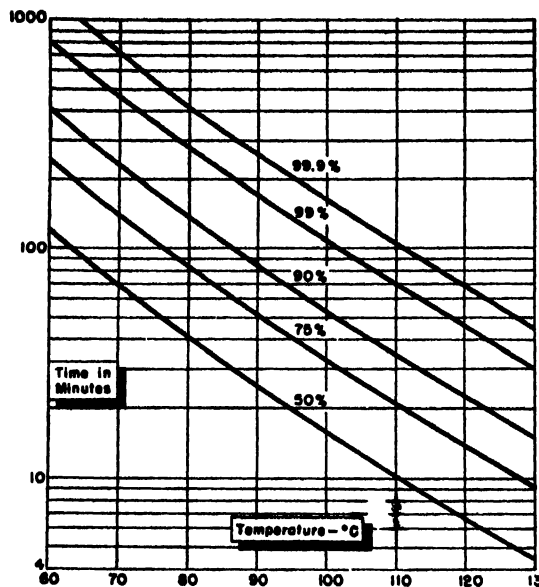


FIG. 5. Effect of temperature on the time necessary for polystyrene to travel 50, 75, 90, 99, and 99.9 percent of its initial distance from equilibrium volume by isothermal expansion or contraction.

cooling at various rates from different annealing temperatures.

It must be realized, of course, that Fig. 5 applies only to an approach to equilibrium by a viscous flow process. Certain molding techniques tend to produce objects with unrelieved stresses due to elongation of polymer chains from their equilibrium lengths, i.e., a certain amount of orientation is "frozen" into the object. This orientation is not objectionable from the standpoint of dimensional stability at room temperature due to the high activation energy for the coiling up of the chains (rubber-like elasticity). However, it does introduce optical inhomogeneity into the material and also limits its use at more elevated temperatures, thus it is frequently desirable to relieve these stresses. A more detailed study of the shrinkage of oriented specimens at various temperatures should provide data for the construction of an analogue of Fig. 5 for the annealing of slightly oriented objects.

SUMMARY

The apparent second-order transition in polystyrene has been demonstrated to be a rate effect.

The equilibrium volume-temperature curve has been determined over the range of from about 20° to 140°C and no indication of a transition found within this interval. Rates of heating or cooling customarily used in thermal expansion measurements result in an apparent second-order transition in the neighborhood of 82°C. The equilibrium cubical thermal expansion coefficient was found to be 4.53×10^{-4} per °C for one sample of polystyrene, but there are indications that this value varies somewhat from one polystyrene sample to another.

At ordinary rates of heating below the apparent transition temperature, or at more rapid rates near and above the transition temperature, polystyrene exhibits a cubical thermal expansion coefficient of about 2.7×10^{-4} per °C, and this value is not dependent on the heating rate as long as it is rapid enough. This fact suggests that two mechanisms operate in the thermal expansion of polystyrene, at markedly different rates, one resulting in almost instantaneous expansion, even at room temperature, and the other being strongly temperature dependent and contributing to the expansion under normal rates of heating only at higher temperatures. The rate of expansion by the second mechanism has been determined at different temperatures, resulting in an energy of activation of 12.0 kcal. This energy of activation would seem to indicate that the second mechanism involves the viscous flow of polymer chain segments rather than the uncoiling of chains as in rubber-like elasticity, which was found to have an activation energy of about 61 kcal. for polystyrene.

Isothermal volume-time curves were found to be exponential in form and this fact, together with the instantaneous expansion coefficient, has enabled us to set up a differential equation defining the volume-temperature curve under non-equilibrium conditions, i.e., at finite heating rates. A comparison was made between an experimental volume-temperature curve and a theoretical curve under the same conditions obtained by graphical solution of the differential equation. An extension of this method of predicting volume-temperature curves to such problems as the expansion of quenched samples and annealing for dimensional stability was indicated.

Abstracts

Following are abstracts of the other papers presented on the program of the second regular meeting, Division of High Polymer Physics of The American Physical Society, Columbia University, January 24-26, 1946.

Thermal Diffusion of Polymers. PETER DEBYE, *Cornell University*.—Experiments on thermal diffusion of high molecular weight polymers in solution were performed using the Clusius-Dickel arrangement. It was found that whereas molecules of small molecular weight have a thermal diffusion coefficient of the order of 1 percent of the ordinary diffusion coefficient, both coefficients are of the same order of magnitude for polymers. It was also found in the special case of polystyrene that the thermal diffusion coefficient increases with increasing molecular weight indicating that thermal diffusion methods can be used for fractionation purposes as well as for the observation of the molecular weight distribution curve. The question whether thermal diffusion of molecules with the same molecular weight depends also on their form (for instance, branched or straight chain) has not yet been answered. The instruments which we have used are cylindrical in form, hot and cold cylinder concentric with a narrow interspace. Other experiments have shown that packing of the interspace may have practical advantages. The thermal diffusion method also seems appropriate for the purification of liquids from small amounts of impurities. The work has been done in collaboration with A. Bueche and was supported by the Rubber Reserve Company.

X-Ray Diffraction by Potassium Laurate Solutions (A Contribution to the Mechanism of Emulsion Polymerization). E. W. HUGHES, *Shell Development Company*.—This paper will describe the results of an x-ray diffraction study, mostly at small scattering angles, of aqueous potassium laurate solutions with and without other added substances. The soap concentration has been varied over the range 10 percent to 50 percent, which includes the phase change from isotropic solution to anisotropic "middle soap." The added materials include KCl up to 2.0 weight normal and more than thirty hydrocarbons and hydrocarbon derivatives. The hydrocarbons have been added in some instances up to a concentration of one mole per mole of soap. These experiments are part of a study of the role of soap in emulsion polymerization.

The Use of Color and Fluorescence Indicators for Determining the Structure of Glasses. W. A. WEYL, *Glass Science, Inc.*.—The fact that light absorption and emission depend on the environment of the ions and molecules which absorb or emit light can be used to draw conclusions about the structure of liquids, glasses, and crystals. This method supplements, to a certain extent, information gained from other sources, such as x-ray diffraction data. The method is illustrated by several examples, like the influence of thermal history on the structure of glass and the constitution of the yellow fluorescing zinc orthosilicate.

The Mechanical Properties of Fibers. HENRY EYRING AND GEORGE HALSEY, *Textile Research Institute Laboratories*.—The stress-strain behavior of textiles has been interpreted quantitatively in terms of a model with a viscous element in parallel with one spring and in series with a second. Such a viscous element is of the non-Newtonian type required in plastic flow. The treatment of more complicated conditions of straining gives rise to equations best treated by the superposing of elementary processes. For any restraint, methods are given for computing the stress-strain curves from the elementary processes.

The Molecular Structure of Some Polyuronides as Revealed by X-Ray Diffraction Studies. K. J. PALMER, *Western Regional Research Laboratory,* Albany, California*.—Recent x-ray diffraction studies carried out on some naturally occurring polymers, particularly polymannuronic acid (alginic acid) and polygalacturonic acid (pectic acid), show that they are essentially long, straight chains. The symmetry of the chain in alginic acid is twofold, in sodium alginate threefold. The polygalacturonide chain remains threefold in pectin, pectinic acid, and pectic acid as well as the Na, K, and Rb pectates. The fiber axis identity period of the polygalacturonide chain is always about 13A. This identity period is compatible with the generally accepted chemical structure for this chain when the pyranose rings are assumed to have one of the two alternative transconfigurations. Moisture isotherms and x-ray data on pectinic acids show that absorbed water hydrates the polar groups and increases the interchain separation. The non-uronide material which occurs to the extent of 10-20 percent in the pectates has no detectable influence on the x-ray pattern.

* Bureau of Agricultural and Industrial Chemistry, Agriculture Research Administration, U. S. Department of Agriculture.

Statistics of Cross-Linked Polymers. WALTER H. STOCKMAYER, *Department of Chemistry, Massachusetts Institute of Technology*.—The present statistical theory of molecular distribution is highly branched and cross-linked polymer will be critically reviewed. The existing theory rests on two assumptions. The first, that all similar functional groups are equally reactive, becomes invalid for dilute solutions or for diffusion-controlled reactions. The second, that no cyclic structures can form, is more serious and must be removed before truly cross-linked polymers can be treated, but the mathematical difficulties are formidable. Experimental observations of gel points will be compared with the theory. For polycondensations, agreement is tolerably good, so that the statistical formulas may be applied with some confidence to these reactions prior to the gel point. Such an application has already been made

to a study of the viscosities of branched polyesters. The theory has now been extended to deal with a general polycondensation reaction involving any number of components. For addition polymers, the theory is generally poor, because its basic assumptions are seriously violated. Finally, the necessity of an improved theory for the gelled state will be emphasized. Flory's treatment, though qualitatively attractive, appears not to be generally applicable. It is concluded that cyclic structures must be specifically considered in the basic formulation of the theory.

Some Thermodynamic Properties of Slightly Cross-Linked Gels. R. F. BOYER AND R. S. SPENCER, *The Dow Chemical Company, Midland, Michigan*.—It has been shown recently¹ that a measurement of the deswelling of slightly cross-linked gels by high polymer solutions affords in principle a new method of determining the number-average molecular weights of polymers. In order to assess the merits of this method, it has been necessary to explore the behavior of gels in some detail. Amount of cross-linking agent, size of the sample, percentage of soluble material present in the gel, variation of swelling ratio in pure solvent for a series of supposedly identical gels, and influence of solvent type were among the factors studied. Measurements of the equilibrium swelling volumes of a series of identical gels in a wide range of solvents afford a convenient method for evaluating solvent power or solvent-polymer interaction. This is done in terms of the parameter μ_0 which arises in the theory of gels. Some fifty solvents have been examined in this regard. Determination of μ_0 for solvent-non-solvent mixtures leads to some interesting results. All work reported here is confined to styrene-divinyl benzene gels.

¹ R. F. Boyer, *J. Chem. Phys.* 13, 363 (1945).

The Determination of Polymer-Liquid Interaction by Swelling Measurements. PAUL DOTY AND HELEN S. ZABLE, *Department of Chemistry, Polytechnic Institute of Brooklyn*.—The interaction of polymers and liquids can be fairly well characterized by the value of the quantity μ , commonly used in thermodynamics of polymer solutions and evaluated from osmotic pressure measurements. The Flory-Rehner theory of swelling relates the amount of swelling of a slightly cross-linked polymer in a liquid to μ and a parameter M_c , the average molecular weight between cross-links. M_c may be determined for a cross-linked polymer from its swelling volume in a solvent in which osmotic pressure measurements with the uncross-linked polymer have given the value of μ . The cross-linked polymer, thus calibrated, can be used to obtain the μ value for other liquids by measuring its swelling volume in them. This procedure has been carried out on polyvinyl chloride which can be cross-linked by heating. Calibration in three different liquids gave comparable values of M_c , thus checking the Flory-Rehner theory. The quantity μ was determined at two different temperatures for fifty liquids. The value of μ and the temperature dependence of μ correlate well with the observed properties of the binary systems. The variations of μ with molecular weight in an homologous series is demonstrated.

Stress-Time-Temperature Relations in Polysulfide Rubbers. M. D. STERN AND A. V. TOBOLSKY, *Princeton University*.—Polysulfide rubbers of various internal structures have been investigated by measurements of continuous and intermittent relaxation of stress and by creep under constant load at temperatures between 35°C and 120°C. Continuous stress relaxation measurements indicate that these rubbers approximately obey the simple Maxwellian law of relaxation of stress, which indicates that one definite type of bond in the network structure is responsible for stress decay. The activation energy for the relaxation process in each of the polysulfide rubbers is nearly the same, indicating that the same type of bond is responsible for the relaxation behavior of all the polysulfides investigated. In contrast to the hydrocarbon rubbers, oxygen is not the cause of high temperature relaxation in polysulfide rubbers, nor does heating in air at moderate temperatures for times comparable to the relaxation time produce permanent chemical changes or changes in modulus. Several possibilities regarding the mechanism of the relaxation process and the type of bond involved are considered in the light of the experimental results.

Viscometric Investigation of Dimethyl Siloxane Polymers. ARTHUR J. BARRY, *Dow-Corning Corporation, Midland, Michigan*.—A series of dimethyl polysiloxane fluids including pure distilled known members of linear and cyclic structures up to ten siloxane units in size as well as a series of intermediate oils of linear structure, whose number average molecular weights have been independently determined, have been studied viscometrically. The intrinsic viscosities were determined and the Staudinger constant evaluated. The molecular weights of several high viscosity polysiloxane fluids were estimated on the basis of this constant and a good fit with the Flory melt viscosity relationship found.

Some Fundamental Relationships between Intrinsic Viscosity, Diffusion and Sedimentation Constants and Thermodynamic Properties of High Polymer Solutions. ROBERT SIMHA, *National Bureau of Standards*.—Empirical relations previously proposed for the variations of the sedimentation and diffusion constants of high polymers are reviewed. If

$$s = s_0/(1 + k_s c), \quad D = D_0(1 + k_D c),$$

then for dilute solutions

$$k_s + k_D = 2bM/RT + v_{sp},$$

where b is the slope of the reduced osmotic pressure curve, M the molecular weight of the polymer, and v_{sp} its specific volume. Since in several instances k_D is small, while in some cases k_s is fairly well represented by the intrinsic viscosity, one may expect in general a certain parallelism between the latter quantity and the thermodynamic term. Analysis of a number of solutions of natural and synthetic polymers over a wide range of molecular weights and of varying slopes b shows the ratio between the thermodynamic factor and the intrinsic viscosity to assume values between one and two for most systems. Barring very poor

solvents, a semi-empirical equation is proposed which relates the intrinsic viscosity to equilibrium properties of the polymer-solvent system, *viz.*

$$(1-2\mu)V_2/V_1 = r[\eta]V$$

is the well-known parameter in the expression for the activities; V_2/V_1 the volume ratio between the two components. $[\eta]V$ represents the intrinsic viscosity, based on volume fractions, and r is a factor which assumes values between one and two in the majority of cases considered.

Thin Section Methods for the Electron Microscopic Examination of Cured Polymers. H. C. O'BRIEN, *Research Laboratories, St. Joseph Lead Company, Josephstown, Pennsylvania.*—A method is described involving the use of the "Cyclone Knife" Ultramicrotome for the preparation of thin sections (0.1 to 1.0 micron) of cured rubber and other high polymer compounds. Such materials which do not permit the formation of film samples for the electron microscope because of their insolubility may thus be examined in thin sections for pigment dispersion and other characteristics. The Ultramicrotome, accessory apparatus, and the preparation and mounting of the specimens for the electron instrument are described. Micrographs of compounded polymers are presented to define the value of the method.

Attachment for Obtaining Angular Distribution of Scattered Light from Measurements with a Ninety-Degree Turbidimeter. J. N. WILSON, *Shell Development Company.*—A device is described which in conjunction with a ninety-degree turbidimeter permits measurement of the angular distribution of the intensity of scattered light over the range 145 to 115 degrees and 65 to 35 degrees. The apparatus consists of two mirrors by means of which the direction of the incident light beam can be varied over the required range. A discussion is given of methods of calibration and of typical results obtained.

A New Approach to the Theory of Relaxing Polymeric Media. M. S. GREEN AND A. V. TOBOLSKY, *Princeton*

University.—A molecular theory of relaxing media is presented which gives an expression for the stress in terms of the strain history. At any given time the strain history produces a distribution in internal strains which for mechanical properties can be characterized by a limited number of internal strain parameters. The second law of thermodynamics is used to define dissipation of energy at constant temperature, and explicit expressions for dissipation of energy for any strain history are obtained. Inasmuch as relaxation during straining causes an essential reorganization of structure which is in fact the cause of dissipation, the kinetic theory of elasticity is extended to non-isotropic polymeric networks. A tensor expression for the stress-strain-time relations is thereby developed.

Application of Molecular Distribution Methods to the Statistical Mechanics of High Polymer Solutions. BRUNO H. ZIMM, *Polytechnic Institute of Brooklyn.*—In the expression for the osmotic pressure of a solution as a power series in the concentration, the term in the second power of the concentration determines the deviations of dilute solutions from ideality. A general expression for the coefficient, A_2 , of this term has been developed for high polymer solutions using the probability distribution function of the configurations of a pair of chain molecules. In favorable cases it is found that the well-known Flory-Huggins equation arises as a first approximation, but that in general serious modifications to this equation must be made. These modifications arise from more complicated interactions of chain segments than are considered in the Flory-Huggins treatment and can be shown both theoretically and experimentally to be very considerable. It can be further shown that A_2 must depend on the molecular size, according to a relation whose first two terms are

$$A_2 = A_2^0(1 - kpA_2^0/R^2),$$

where A_2^0 and k are constants, p is the degree of polymerization, and R is the root-mean-square distance between chain ends.

Here and There

New Appointments

Twelve new staff members have been appointed to the University of Chicago's Institute of Nuclear Studies. They are Willard F. Libby, full professor, and the following assistant professors: Herbert L. Anderson, Nicholas C. Metropolis, Nathan Sugarman, Anthony L. Turkevich, James S. Allen, Stanley P. Frankel, Eldred C. Nelson, Harrison S. Brown, Norman S. Elliott, T. Harrison Davies, and Clyde A. Hutchison.

Robert B. Jacobs, until recently an engineer for the Kellogg Corporation, has joined the staff of Distillation Products, Inc., Rochester, New York, where he will direct the work of the physics laboratory in the study of physical phenomena at low pressures.

William A. Shurcliff, Scientific and Technical Adviser to the New York State Department, has been given temporary leave of absence by Governor Thomas E. Dewey to serve as official historian of "Operation Crossroads," to be carried out in the Marshall Islands in May and June by Joint Task Force One. Dr. Shurcliff helped edit the *Smyth Report* on the atomic bomb.

Noel C. Jamison joined Philips Laboratories, Irvington, New York, on January 2 as division chief in charge of electro-acoustics.

Charles T. Zavales recently joined the medical x-ray division of North American Philips Company as technical-commercial adviser.

Awards

Captain Ralph Decker Bennett, technical director of the Naval Ordnance Laboratory, has been awarded the Legion of Merit for his contributions in Naval ordnance research and development projects. Captain Bennett came to the Laboratory in 1940 from Massachusetts Institute of Technology, where he was professor of electrical measurements.

Enrico Fermi, Harold C. Urey, Samuel K. Allison, Cyril Smith, and Robert S. Serber, all of the University of Chicago, were presented Medal of Merit citations for their central roles in the production of the atomic bomb on March 20. Major General Leslie R. Groves, chief of the Army's Manhattan District, made the awards.

American Telephone and Telegraph Company has announced the award of five Frank B. Jewett fellowships for research in the physical sciences. Recipients are: Martin G. Ettlinger, Edward W. Fager, Bernard Goodrich, Shuichi Kusaka, and Robert L. Scott.

Reinhold Rudenberg has received the Honor Award of Stevens Institute of Technology "for notable achievement in the field of electron optics as the inventor of the electron microscope."

Ohmite Laboratory

Ohmite Laboratory at Illinois Institute of Technology, established to provide service to industry in the field of electrical measurements, began in February to make its facilities available to sponsoring organizations. The result of a \$32,500 contribution by David T. Siegel, president of the Ohmite Manufacturing Company, this laboratory will provide precision measurement of electrical and magnetic quantities for the Chicago area approaching in accuracy those of the Bureau of Standards in Washington. It is jointly operated by Illinois Institute's Electrical Engineering Department and the Electrical Engineering Division of the Institute's Armour Research Foundation.

Congress for Applied Mechanics

The Sixth International Congress for Applied Mechanics will be held in Paris from September 22 to 29, 1946. It will be divided into the following sections:

1. Structures. Elasticity. Plasticity.
2. Hydro- and Aerodynamics. Hydraulics.
3. Solid Dynamics. Vibration and Sound. Friction and Lubrication.
4. Thermodynamics. Heat Transfer. Combustion. Fundamentals of Nuclear Energy.

Those who desire to become members of the Congress are requested to inform the Secretary General as soon as possible of their intention to attend the Congress. They should also indicate whether they wish to present a paper, and in what section. For information address the Secretary General at Institut Henri-Poincare, 11, rue Pierre-Curie, Paris (V*), or write to Dr. Josephine de Karman, 1501 South Marengo Avenue, Pasadena, California.

Westinghouse Forum

A science and engineering forum honoring the centennial of the birth of George Westinghouse will be held by Westinghouse Electric Corporation in Pittsburgh on May 16-18. Internationally prominent scientists will be among the speakers.

Products of Tomorrow Exposition Postponed

Because of the uncertainty of products and delivery schedules of large numbers of the nation's leading manufacturers, the Products of Tomorrow Exposition scheduled to open at the Chicago Coliseum April 27 has been indefinitely postponed. It may be held by the fall of this year or held over to early in 1947.

New Books

What Are Cosmic Rays

By PIERRE AUGER. Translated from the French by Maurice M. Shapiro. Pp. 128, Plates 22. University of Chicago Press, 1944.

During the last fifteen years the subject of cosmic rays has received a great deal of attention from both experimental and theoretical physicists. A great many articles have appeared in the technical literature. On the other

hand, there have been practically no books written by authoritative individuals, describing the progress made in this subject. It is therefore a particular pleasure to welcome a book by Pierre Auger. Dr. Auger is well known for his investigations and for his excellent experimental techniques, and the book therefore carries the weight of a distinguished investigator in the subject. The book was translated from the French by Dr. Shapiro who also has done experimental work in the field of cosmic rays.

The book starts with an account of the initial discovery of the residual ionization which led to the recognition that cosmic rays exist. From this beginning the development of the subject is traced through the picturesque phase of the worldwide survey expeditions which established the particle nature and energy spectrum of the radiation. The bulk of the book is devoted to the present stage of research in which most of the work is done by rather complex arrangements of Geiger counters and cloud chambers and is sometimes further implemented by high altitude balloon or airplane flights. Dr. Auger describes the experiments leading to the recognition of the positive electron and the positive and negative mesotron. The author also describes many of the more recent experiments on "cascade" shower production, knock-on electrons, mesotrons, and nuclear explosion processes. He discusses the evidence which tends to show that the primary rays are protons. This reviewer only regrets that the book was not longer.

In the preface, the translator says: "This book is written primarily for the reader who lacks a technical knowledge of physics but who wants to keep in touch with current developments in science. It is not intended for the specialist but should provide a good perspective for the physicist who wants to get a quick view of what has been done in this new field." The book is written on a semi-technical level, without recourse to mathematics, and presupposes some knowledge of the main phenomena involved. It is especially suitable for graduate students and professional physicists whose main interest is in other branches of physics, as well as for research workers in allied sciences who are familiar with physical terminology. For these persons it presents a bird's eye view of present conceptions of the subject.

The binding is of good quality, the type is easy to read, and the photographs at the end which illustrate the main cosmic-ray phenomena are both interesting and instructive.

SERGE A. KORFF
New York University

Lehrbuch der Funktionentheorie

By L. BIEBERBACH. Two volume set. Vol. I. *Elemente der Funktionentheorie*. Fourth edition. Pp. 322+xiv, Figs. 77, $5\frac{1}{2} \times 8\frac{1}{2}$ in. Vol. II. *Moderne Funktionentheorie*. Second edition. Pp. 370+vi, Figs. 47, $5\frac{1}{2} \times 8\frac{1}{2}$ in. Chelsea Publishing Company, New York, 1945. Price for set \$6.50.

These treatises are reprints of the latest German editions. The latest edition of Vol. I was originally published by J. Springer, Berlin, in 1934; Vol. II, in 1931.

Volume I. The first edition of this volume was published

25 years ago; yet the fourth edition still is one of the finest introductions to the theory of functions of a complex variable. The treatment, though elementary, is quite complete. Full use of the Cauchy, Weierstrass, and Riemann points of view have been made. The talent of the author in combining a clear presentation with a reasonably comprehensive treatment of the subject is almost everywhere in evidence. The reader will be quick to feel the author's deep interest in the field.

The following chapter headings give an indication of the scope of Volume I: complex numbers, series, functions of a complex variable, special functions, conformal mapping, complex integrals, Cauchy integral formula, residues, analytic continuation, algebraic functions, integrals of algebraic functions, simple periodic functions, elliptic functions, representation of analytic functions by series and products, gamma-function.

Volume II. Volume II contains a great variety of material on modern theory of functions of a complex variable. In it are reflected the many advances of the subject up to the time of its original publication.

This volume contains extensive treatments of such topics as the following: conformal mapping, elliptic modular functions, theory of analytic functions bounded in the unit circle, Professor Bieberbach's treatment of uniformization, Picard's theorem, entire functions, analytic continuation, zeta-functions.

These classical treatises will undoubtedly continue to be useful for many years to come, to those whose primary interest is in mathematics as well as to those whose concern lies mainly in the physical sciences. Serious students of physics, engineering, and related fields who make frequent use of functions of a complex variable, conformal mapping, and the like, will profit by a thorough study of these volumes.

RICHARD S. BURINGTON
Washington, D. C.

Physical Methods of Organic Chemistry, Volume I

By ARNOLD WEISSBERGER, Editor. Interscience Publishers, Inc., New York, 1945.

In the preface to this volume, Arnold Weissberger suggests that this work has been compiled to assist the chemist in acquainting himself with physical methods and the student in understanding the principles involved in the methods described. In addition, the book should also be of value to still a third person, the engineer who desires information as to the accuracy of data and what information may be reasonably requested from the laboratory.

Beginning with the "Determination of Melting and Freezing Temperatures," the successive chapters describe the theory and apparatus necessary for measuring boiling point, density, solubility, viscosity, interfacial tension and parachor, osmotic pressure and diffusivity of pure substances and solutions. While the needs of organic chemistry are particularly stressed, the applications and principles involved need not be confined to any one field. Quite naturally, the examples used in describing operations are those of organic chemicals.

Wherever possible, the authors have included alternate methods of determination for various required degrees of accuracy or for the different types of materials to be measured. Very useful are tables, such as that on page 105 in which means of determining density are recommended in terms of the nature of the substance, the conditions required (such as speed, size of sample, etc.), and the maximum precision desired.

Particularly noticeable is the recommended use of comparative rather than absolute methods of measurement. Through this technique the sources of error inherent in many laboratory methods may be minimized. Frequently the property of a reference substance has been established as standard or can be determined very accurately by a difficult or complex method. In this case, the comparison determination may be made much more rapidly and in simpler apparatus than if an absolute determination were necessary.

Other chapters discuss the properties of monolayers and duplex films, calorimetry, microscopy, crystal form and crystallochemical analysis, diffraction (both x-ray and electronic), and refractometry.

Many references are given, including specific articles, listed by number on the same page as the corresponding item in the text, and general references, which are tabulated at the end of each chapter. Thus, a sound basis is provided for anyone who might wish to pursue any particular subject farther. However, sufficient information on formulas, constants, and standards has been included so that use of the references is not necessary and may be made at the discretion of the reader.

Volume I has been published without an index. While not absolutely necessary, as the Table of Contents and the Chapter Outlines enable any subject to be located fairly readily, an index would be a valuable addition to any book of this nature. Perhaps one is included in Volume II.

As might be expected in a work of this nature, the style of presentation varies throughout the book. This is no disadvantage as each chapter is complete in itself. Both the authors and the editor are to be commended for maintaining the uniform high quality of the work throughout the diverse subjects which are included.

THOMAS J. WALSH
The Standard Oil Company (Ohio)

New Booklets

National Research Corporation, 100 Brookline Avenue, Boston, Massachusetts, has issued a 30-page book, containing many pictures, entitled *High Vacuum for Industry*, describing the new commercial possibilities for high vacuum.

Interchemical Review for Autumn 1945, published by the Interchemical Corporation, 432 West 45 Street, New York 19, New York, has the following table of contents:

Analysis by Excited Atoms
Colors for Textiles: Ancient and Modern
Proximity Radio Fuse
Glass Blank for War Planes

Fischer and Porter Company, Hatboro, Pennsylvania, has issued an 8-page bulletin explaining Rotameter advantages and covering the most prominent styles of rotameters and their uses. Free. Request Bulletin 45-A from Fischer and Porter Company, Department 5-4, County Line Road, Hatboro, Pennsylvania.

Alemite Division of Stewart-Warner Corporation, Chicago, Illinois, has published a book entitled *Alemite Answers*, containing a discussion of man's age-long struggle with friction and a graphic presentation to demonstrate, in laymen's pictures and terms, the how, what, when, where, and why of lubrication. The book is for distribution at \$3 per copy through the Alemite distributor organization, to industry and the automotive, marine, and aviation fields. In addition, copies will be made available to engineering and technical schools, designers, engineering societies, and other individuals and groups concerned with machinery.

Eitel-McCullough, Inc., San Bruno, California, has issued a 20-page brochure, fully illustrated, which presents the entire Eimar line of electronic products. Free on request.

Bakelite Corporation, 30 East 42 Street, New York 17, New York, publishes quarterly the *Bakelite Review*, the purpose of which is "to inform industry of the latest developments in plastics and synthetic resins and their applications." The January, 1946 issue contains 32 pages and many attractive illustrations.

Science and Appliance for January, 1946, a four-page folder issued monthly by The Ohio State University Research Foundation, Columbus 10, Ohio, features articles on the work of A. R. Olpin, until recently the Executive Director of the Foundation: Sonar—underwater radar; progress in aids to hearing; uses of glass fibers.

Allied Radio Corporation, 833 West Jackson Boulevard, Chicago 7, Illinois, has announced the publication of a new 1946 buying guide, *Everything in Radio and Electronics*. Free on request.

Ohmite Manufacturing Company, 4835 West Flournoy Street, Chicago 44, Illinois, has issued a four-page folder, Bulletin No. 126, on Riteohm Precision Resistors. Free on request.

Philco Corporation, Philadelphia 34, Pennsylvania, is sending to all Philco stockholders, distributors, and dealers a 32-page booklet, *Radar on Wings*. "This first complete story of airborne radar is illustrated with a large number of photographs and drawings of operational and installation views that were stamped 'Secret' until long after V-J Day and can only now be released."

Bell Telephone Laboratories, Inc., 463 West Street, New York 14, New York, has published a 28-page booklet entitled *Radar and Your Telephones*. It tells in non-technical language how the problems of radar were solved and how it accomplished so much in actual combat. Many color illustrations.

Journal of Applied Physics

Volume 17, Number 6

June, 1946

A Study of Distortion in Electron Microscope Projection Lenses*

JAMES HILLIER

RCA Laboratories, Princeton, New Jersey

(Received January 18, 1946)

The origin of distortion in electron microscope projection lenses is discussed and the serious nature of its effect on the measurement of particle size distributions is pointed out. Methods of measuring distortion are described. By means of first-order theory it is shown to be possible to correct distortion by the use of a two element projection lens. The degree of correction obtainable is shown to be satisfactory for most practical purposes. A double gap projection lens polepiece and the correction of distortion obtained with it are described.

I. INTRODUCTION

IN light optics it is possible to minimize the aberrations of a particular system by the appropriate combination of simple lenses. In precision electron optics, on the other hand, this has not yet been accomplished. Fortunately, while the aberrations of a simple electron lens are of the same order as those of the corresponding light optical lens, it is possible to take advantage of the very short wave-length associated with the electron beam and minimize the aberrations by radically reducing the aperture of the lens. Thus, even with a simple electron lens it is possible to produce high quality images with resolving powers of approximately 1×10^{-7} cm.

An electron microscope capable of this resolving power more than satisfies most of the immediate needs of workers interested in the fine structure of matter beyond that observable with the light microscope. This has resulted in the

more concentrated efforts of electron microscope designers being directed towards the development of a practical instrument rather than to the improvement of the lens system. The fact that, in the case of the electron microscope projection lens, the aberrations produce secondary effects of a surprisingly high order of magnitude seems to have been overlooked in the literature. Some of these effects and means of correction will be discussed in the following.

Image formation by the projection lens differs in many ways from that by the objective lens. In the objective a region of the specimen, small compared with the area of the lens utilized by a single pencil,¹ is imaged by pencils of electrons having an angular aperture in the neighborhood of 3×10^{-3} radian. All points of the image are diffused by the spherical and chromatic aberrations while the off-axis points are displaced and further diffused by the additive effects of other third-order aberrations. Actually, because the

* Presented before the Electron Microscope Society Meeting in Princeton, New Jersey, November 30, 1945.

¹ The term "pencil" is used to designate those rays which take part in the imaging of a single point.

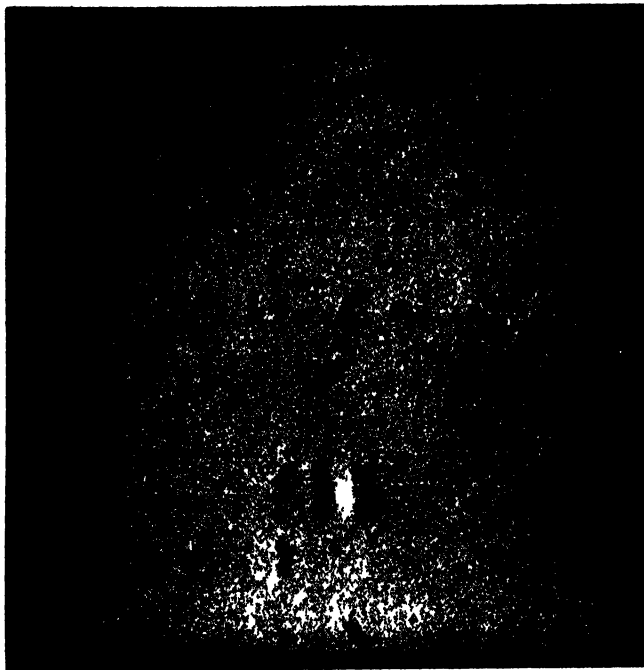


FIG. 1. Electron micrograph of a dispersion of zinc oxide pigment. In its extreme corners this image has a radial distortion (D_r) of about 30 percent. $\times 12,000$.

diameter of the field of view is small, the last are almost negligible.² The image produced by the objective constitutes the "object" for the projection lens. The area of the object, in this case, is large compared with the area of the projection lens utilized by a single pencil. As the individual pencils utilized in the projection lens have an extremely small angular aperture (10^{-5} rad.) none of the third-order aberrations appears as an observable diffusion of the image. In other words, *the aberrations of the projection lens cannot reduce the resolving power of the instrument.*³ Instead the aberrations appear as a displacement of the image points from the Gaussian or true positions; that is, as a distortion. Actually, when the projector is considered as a thin lens the conditions of imaging are such that the spherical aberration

² See V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945).

³ The combined effects of chromatic change in magnification and chromatic change in rotation (in the case of a magnetic lens) may be considered as an exception to the above statement. However, these affect only the outer zones of the image and are less, in general, than the defocusing of the entire image caused by the simultaneously appearing chromatic aberrations of the objective.

alone is responsible for this distortion. In the case of a magnetic lens there is also rotational distortion of the image but this is usually slight.

In addition to the fact that distortion in electron microscope images is undesirable from an aesthetic point of view, almost undetectable amounts can produce difficulties in the production of good stereoscopic images and can lead to completely erroneous results if the microscope is being used for the accurate determination of particle size distributions. Figure 1 is an electron micrograph of a dispersion of fine particles. This image is known to possess pincushion distortion to the extent that the radial magnification is 30 percent greater at the edge than at the center; yet it appears quite normal to the eye. Figure 2 is an electron micrograph of a regular wire screen showing the same amount of distortion. The fact that the image is distorted is apparent in this figure, though it does not appear to be excessive.

The effect of this amount of distortion on a particle size distribution is illustrated in Fig. 3. Curve A represents the particle size distribution which would be obtained from an undistorted

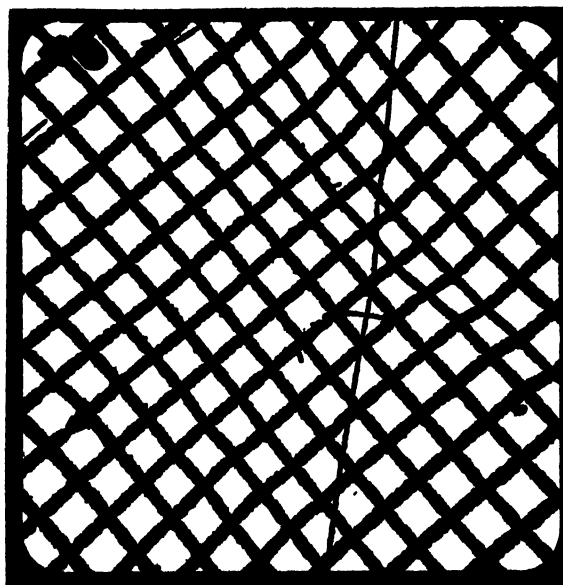


FIG. 2. A micrograph of a fairly regular wire screen made with the same projection lens and under the same conditions of distortion as the micrograph in Fig. 1.

image of a uniform dispersion of particles which are all exactly 1000Å in diameters. It is assumed that the instrument is working at a resolving power of approximately 25Å. Curves B and C, on the other hand, represent the extreme distributions which would be obtained from measuring electron micrographs in which the above amount of distortion is present. Curve B would be obtained when particles are all measured in the circumferential direction in the image and Curve C when the particles are measured in the radial direction. In practice the distribution obtained would most likely lie between these two extremes, the actual curve depending on the particular arrangement used in the measurement of the images of the particles. *Thus a relatively slight amount of distortion in the electron microscope image produces a serious error in the accurate determination of a particle size distribution.*

In order to correct for such errors at present, it is necessary for the electron microscopist to calibrate the image field accurately before

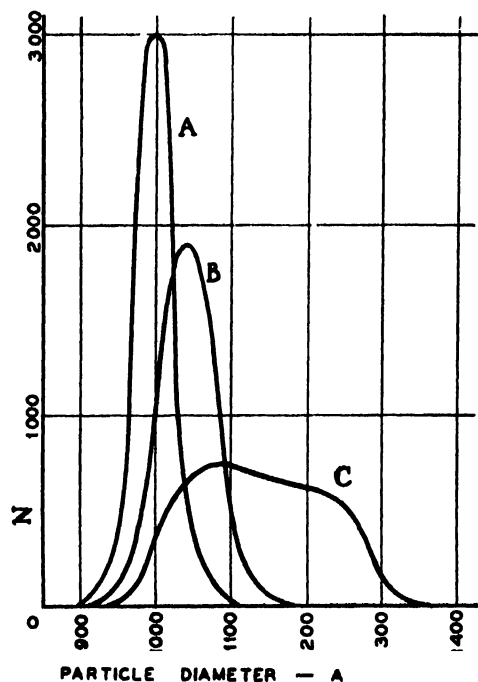


FIG. 3. A group of curves showing the effect on a particle size distribution produced by distortion of the amount present in Figs. 1 and 2.

- A. Particle size distribution of a uniform suspension of particles exactly 1000Å in diameter obtained with a distortion-free system.
- B. Particle size distribution of the same specimen which would be obtained by measurement in the circumferential direction in an image possessing distortion ($D_r = 30$ percent at edge).
- C. Particle size distribution of the same specimen which would be obtained by measurement in the radial direction.

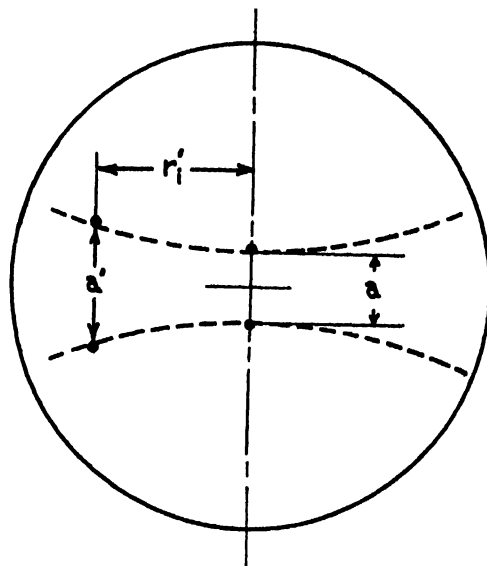


FIG. 4. Diagram showing the measurements which permit the complete determination of pincushion distortion at any point of an image.

making critical particle size determinations. Of greater importance and convenience is the correction of the distortion by proper design of the projection lens.

II. MEASUREMENT OF DISTORTION

Distortion in an electron microscope image involves a displacement of the image points, usually in the radial direction, from a geometrically true representation of the specimen. Thus, if r_i' , θ_i' are coordinates of the actual image point and r_i , θ_i , the coordinates of the corresponding point in the geometrically true image

$$r_i' = F(r_i); \quad \theta_i' = \theta_i + G(r_i), \quad (1)$$

where $F(r_i)$ and $G(r_i)$ are continuous, single-valued functions of r_i . $G(r_i) = 0$ except in the case of the magnetic projection lens; in any case it is small and will be neglected in the present work. The expressions (1) represent the transformation of coordinates necessary to transform the true image into the distorted image. In practice, however, the distortion can be considered as a point to point variation in the magnification since that is the way in which it disturbs the accuracy of measurements made in the image. It is obvious that the variation in

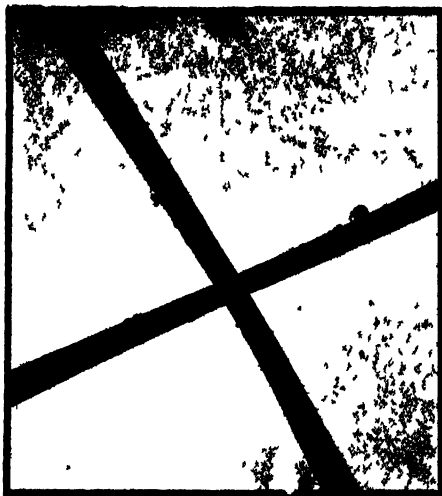


FIG. 5. An image produced by the projection lens of stretched cross wires from which it is possible to determine completely the distortion pattern of the lens.

magnification depends on the direction of measurement in the image.

Consider now two points of the image which are separated by a distance a and which straddle the central point of the image (see Fig. 4). By moving the specimen in such a way that these points move away from the center of the image in the direction perpendicular to the line joining them they will trace out two curves which, as will be seen below, give all the information necessary for the complete determination of the distortion.

It is immediately obvious that $(2\pi r_1'/a') = (2\pi r_1/a)$ and hence the circumferential distortion D_c is given by the expression:

$$(D_c/100) + 1 = a'/a = r_1'/r_1 = F(r_1)/r_1, \quad (2)$$

where D_c is defined as the percentage difference between the magnification at the point under consideration (measured in the circumferential direction) and that at the center of the image. For high accuracy a must be small.

The radial distortion D_r , which is similarly defined with the exception that the measurements are made in the radial direction is given by the expression:

$$\frac{D_r}{100} + 1 = \lim_{r_{12} - r_{11} \rightarrow 0} \left(\frac{r_{12}' - r_{11}'}{r_{12} - r_{11}} \right) = F'(r_1). \quad (3)$$

The actual procedure of measuring the dis-

tortion is now quite straightforward. The necessary data are obtained by making a series of micrographs in which a selected region of the specimen is moved in the way indicated in Fig. 4. If the projection lens of the instrument is easily accessible, it is more convenient to arrange two cross-wires in a holder just above it and make a single micrograph as shown in Fig. 5. Ordinary finely drawn wires can be depended on to have the same diameter over a considerable length. In either case it is possible to obtain precise measurements of (a'/a) and r_1' from which a curve can be drawn showing the relationship between D_c and r_1' . From these measurements it is also possible to calculate the corresponding values of r_1 and plot r_1' as a function of r_1 . The value of $[(D_r/100) + 1]$ can then be determined for any point of the image by taking the slope of the curve for r_1' as a function of r_1 at the desired points and plotting against r_1' . Figure 6 gives two curves for a typical projection lens system working at a low magnification for which the excessive distortion at the edge of the image is visible. From such measurements a calibration chart can be drawn up which gives the radial and

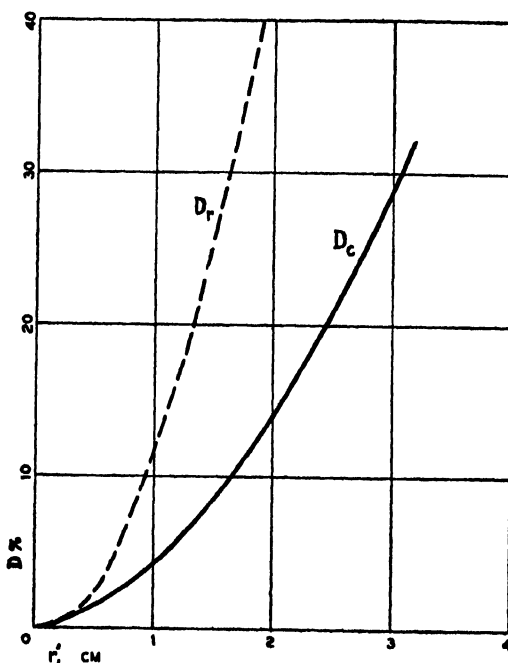


FIG. 6. Radial (D_r) and circumferential (D_c) distortion curves for an early projection lens polepiece working at a low magnification for which the excessive distortion at the edges of the image is visible.

circumferential distortion for each zone of the electron microscope image. If, in particle size measurements, a standard electronic magnification, optical magnification, and measuring technique is used, it is a simple matter to utilize such a chart to compensate for the effects of image distortion. It is important, of course, that the electron-optical center of the projection lens image be accurately known, if true correction is to be obtained. This is normally determined at the same time as the calibrating measurements are made.

III. CORRECTION OF DISTORTION IN A SINGLE LENS

Incidental to the study of distortion of magnetic projection lenses, it was observed that the distortion is dependent on the spacing of the polepieces, being reduced as the spacing is increased. The effect is also predicted from a theoretical study of spherical aberration.² The experimental curve shown in Fig. 7 illustrates this effect. While the improvement attainable is appreciable it is accompanied by considerable loss of magnification. The method is of some practical and academic interest, but it does not represent a complete solution of the problem.

IV. TWO-LENS CORRECTION OF DISTORTION

It was realized a few years ago that the use of a double projection lens with the first lens situated *between the second lens and its object side focal point* provided some correction of image distortion if the strength of the two lenses were carefully adjusted. The qualitative idea was tested and the results described at the time,⁴ but it was not investigated further because of the apparent necessity of using a specially constructed double projection lens coil and an extra regulated current supply. Figure 8 shows the initial results obtained while Fig. 9 is a ray diagram explaining the correcting action of the second lens.

In the present work a mathematical study of such a system has been carried out to show that it is possible to obtain any desired degree of correction with a properly adjusted two lens system

⁴ J. Hillier and R. F. Baker, "Miscellaneous research on electron microscope parameters," presented before Electron Microscope Society of America, January 14, 1944.

and that in practice it is possible to design a distortion-free polepiece system actuated by a single lens coil.

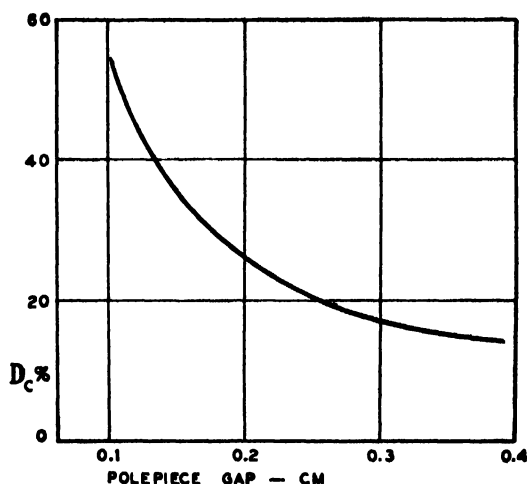


FIG. 7. Curve showing the effect on circumferential distortion of increasing polepiece gap. The curve was taken at constant excitation of the lens, and hence there is an accompanying change in magnification. By drawing the curve for a fixed position in the intermediate image, the effect of this change in magnification is eliminated.

V. THEORY OF TWO-LENS CORRECTION

Figure 10 is a ray diagram indicating the significance of the symbols used in the following. The system consists of two thin lenses of paraxial focal lengths f_{10} and f_{20} separated by a distance d_1 . The image produced by the objective and remagnified by these lenses consists of pencils essentially parallel to the axis and of extremely small angular aperture (10^{-5} rad.). For all practical purposes these pencils can be considered as single rays leaving a distant point.⁵

Since pincushion distortion only is being considered the problem involves determining the characteristics of the second lens which will make r_1' a linear function of r_1 . From the geometry of Fig. 10 it can be seen that

$$r_2 = r_1(f_1 - d_1)/f_1, \quad (4)$$

$$r_1' = r_2(d - d_1 - v_2)/v_2. \quad (5)$$

Then if the simple lens formula $(1/v) - (1/u) = (1/f)$ is assumed to hold for both lenses (5)

⁵ For simplicity this point will be considered at infinity in the present paper. In critical work, however, it is necessary to take into account the fact that the objective is normally at a finite distance from the projector in which case u_1 is substituted for f_1 in the following formulas.



FIG. 8. Projection lens images of parallel stretched wires showing correction of distortion obtained in early work on two-lens projection systems. (a) Uncorrected. (b) Corrected.

becomes, after multiplying and collecting terms:

$$r_i' = r_1 \left[(d-d_1) \left(\frac{1}{f_1} + \frac{1}{f_2} \right) - \frac{d_1(d-d_1)}{f_1 f_2} + \frac{d_1}{f_1} - 1 \right]. \quad (6)$$

Because of spherical aberration f_1 and f_2 depend on r_1 and r_2 , respectively, according to the relations

$$f_1 = f_{10}(1 - k_1 r_1^2) \quad \text{and} \quad f_2 = f_{20}(1 - k_2 r_2^2), \quad (7)$$

where k_1 and k_2 are the spherical aberration coefficients.⁶

Substituting the expressions (7) for f_1 and f_2 in (6) leads to

$$r_i' = r_1 \left[(d-d_1) \left\{ \frac{1+k_1 r_1^2}{f_{10}} + \frac{1+k_2 r_2^2}{f_{20}} - \frac{d_1(d-d_1)}{f_{10} f_{20}} - (1+k_1 r_1^2 + k_2 r_2^2) + \frac{d_1}{f_{10}} (1+k_1 r_1^2) - 1 \right\} \right], \quad (8)$$

where $k_1 r_1^2$ and $k_2 r_2^2$ are considered small compared to 1 in view of the experimental values obtained for these quantities. Collecting coefficients of $k_1 r_1^2$ and $k_2 r_2^2$ results in an expression of the form

$$r_i' = r_1 (A_0 + A_1 k_1 r_1^2 + A_2 k_2 r_2^2). \quad (9)$$

Hence to make r_i' a linear function of r_1 , it is necessary to equate the coefficients of the higher

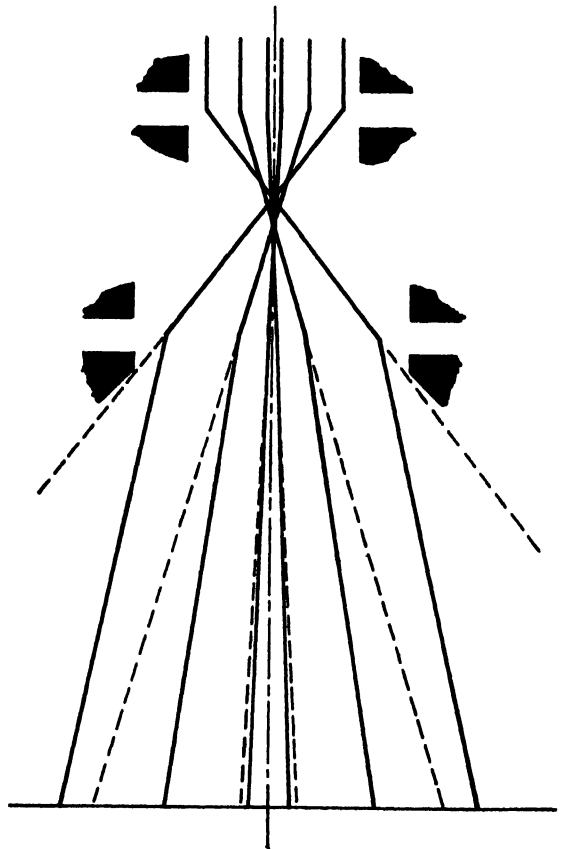


FIG. 9. Ray diagram showing how the combined effects of focusing and spherical aberration in a weak additional component of projection lens system can correct for pincushion distortion. The focal point of the lower lens is situated some distance above the plane of the upper lens.

⁶ The constant k defined above is more convenient for discussing spherical aberration in electron lenses than the commonly used spherical aberration coefficient C . C is a dimensionless constant for a lens of fixed focal length. However, one of the niceties of electron optics is the fact that the focal length of a lens is variable over wide ranges, in which case the use of C becomes cumbersome and leads to difficulties in the interpretation of experimental results. The constant k , on the other hand, is independent of the focal length provided that the lens in question may be regarded as a thin lens. The two coefficients are related by the equation $k f^2 = C f$.

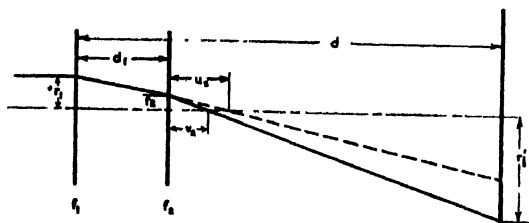


FIG. 10. Ray diagram showing the significance of the symbols used in the text.

powers of r_1 and r_2 to zero, that is

$$A_1 = \frac{d-d_1}{f_{10}} - \frac{d_1(d-d_1)}{f_{10}f_{20}} + \frac{d_1}{f_{10}} = 0, \quad (10)$$

$$A_2 = \frac{d-d_1}{f_{20}} - \frac{d_1(d-d_1)}{f_{10}f_{20}} = 0. \quad (11)$$

Solving (10) and (11) for f_{10} and f_{20} gives

$$f_{10} = d_1, \quad (12)$$

$$f_{20} = -\frac{d_1(d-d_1)}{d} = -\frac{f_{10}(d-f_{10})}{d}. \quad (13)$$

Thus the above has shown that a first-order correction of distortion can be made by the use of two lenses where the second lens is placed at the focal point of the first lens and has a focal length given by (13). Since $f_{10} \ll d$ the focal lengths of the two lenses are very nearly equal. It should also be noted that the conditions for correction are independent of the spherical aberration of either lens.

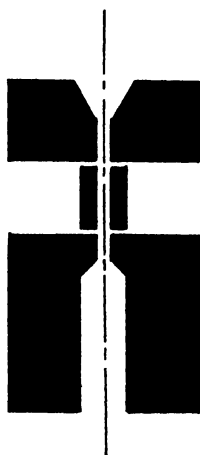


FIG. 11. The schematic cross section of a double gap projection lens polepiece which was used in the present experiments.

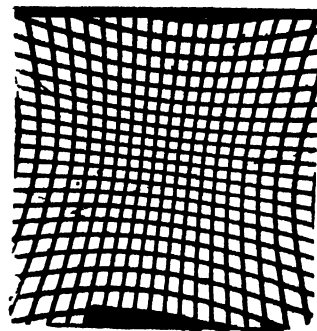


FIG. 12a. The distortion pattern obtained with a single section of the double polepiece section activated.

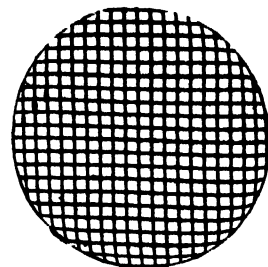


FIG. 12b. Distortion pattern obtained with both sections of the double lens polepiece activated.

To investigate the degree of correction attained by the above method it is necessary to return to the complete expression (5) which on substituting for u_2 and making use of (4) becomes

$$\frac{r_1'}{r_1} + 1 = \frac{d-d_1}{f_2} \cdot \frac{r_2}{r_1} + \frac{d}{f_1}. \quad (14)$$

In order to simplify the notation

$$(r_2/r_1) = (f_1 - d_1)/f_1$$

will be written as ϵ . Then (14) may be written as

$$\left(\frac{D_c}{100} + 1\right)M + 1 = -\frac{(d-d_1)\epsilon}{f_2} - \frac{d(1-\epsilon)}{d_1}, \quad (15)$$

where $M = (r_i/r_1)$ is the undistorted magnification of the system. If the values for the corrected system are introduced it is also possible to write

$$r_1^2 = -\epsilon/[k_1(1-\epsilon)] \quad (16)$$

and

$$r_2^2 = -\epsilon^3/[k_1(1-\epsilon)], \quad (17)$$

whereupon the complete expression for f_2 becomes

$$f_2 = -\frac{d_1(d-d_1)}{d} \left[1 + \frac{k_2\epsilon^3}{k_1(1-\epsilon)} \right]. \quad (18)$$

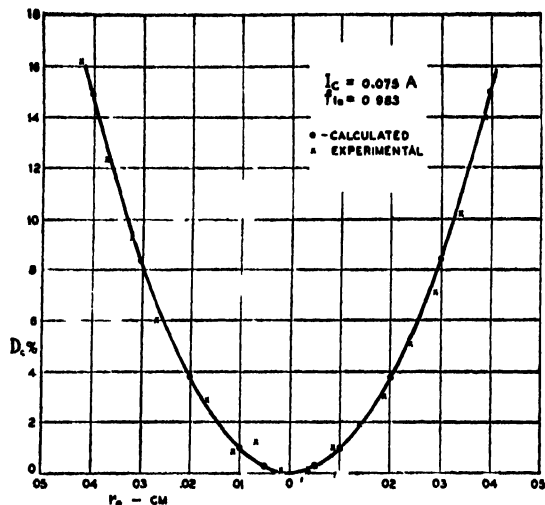


FIG. 13. A comparison between the calculated and experimental distortion pattern for a single section of the polepiece shown in Fig. 11.

Substituting (18) in (15) and rearranging gives

$$\frac{D_0}{100} = - \left(1 + \frac{1}{M} \right) \times \left[\frac{k_2}{k_1} \frac{\epsilon^4}{1 - \epsilon} / \left(1 + \frac{k_2}{k_1} \frac{\epsilon^3}{1 - \epsilon} \right) \right]. \quad (19)$$

Since $\epsilon = -k_1 r_1^2 / (1 - k_1 r_1^2)$ and $k_1 r_1^2 \ll 1$ the expression (19) shows that a system corrected by the application of the relations (12) and (13) provides an image which deviates from the true image only in the fourth power of the small quantity $k_1 r_1^2$.

Before leaving the theoretical discussion, it should be pointed out that even more complete correction can be obtained under the present assumptions if only the condition (12) is imposed on (15) and the expression

$$\left(\frac{D_0}{100} + 1 \right) M + 1 = \frac{d}{d_1} + \epsilon \left[\frac{d - d_1}{f_{20} \left[1 + \frac{k_2}{k_1} \frac{\epsilon^3}{1 - \epsilon} \right]} - \frac{d}{d_1} \right]. \quad (20)$$

In this case a more favorable value of f_{20} can be obtained by making the coefficient of ϵ equal to zero for ϵ equal to its maximum possible value $\epsilon_{\max}$ in a given practical example, that is by

taking

$$f_{20} = \frac{d_1(d - d_1)}{d \left[1 + \frac{k_2}{k_1} \frac{\epsilon_{\max}^3}{1 - \epsilon_{\max}} \right]}. \quad (21)$$

VI. EXPERIMENTAL RESULTS

The realization in practice of a projection lens system which fulfills the above derived conditions is not difficult if it is possible to excite the individual lenses independently. Unfortunately, this requires a rather elaborate special coil structure and an additional regulated source of current if it is to be easily adaptable to existing electron microscopes. On the other hand, if the two lenses are made part of one magnetic circuit excited by a single coil, the proper ratio between the focal lengths of the lenses must be attained by the exact adjustment of the physical shapes and dimensions of the lenses. Owing to the lack of accurate methods of computing the relative field strengths in the lenses of such an arrangement that adjustment can be accomplished only by tedious empirical methods. For these reasons a direct experimental test of the above theory was not attempted in the present work. Instead, a simple special case which could easily be achieved in practice and to which the essential formulas could be applied was selected.

An experimental polepiece assembly was constructed and tested in a demountable electron-optical system. This assembly, which is shown schematically in Fig. 11, included two mechanically identical magnetic lenses separated by 0.9 cm, giving maximum correction of distortion under conditions corresponding to a magnification of approximately $4000\times$ in an RCA Type B electron microscope. The flat-face polepiece shape for the two lenses was chosen because it was known from previous work that such a shape gives particularly high distortion.

As a first test, the lower lens was inactivated and the distortion pattern of the upper lens obtained for a series of different excitations. Figure 12a is one of the patterns obtained. An accurately made electrolytic copper screen (500 mesh) was used to determine the distortion patterns of the system. In Fig. 13 the circumferential distortion D , is plotted against the radial position in the object (r_0). In the

experimental arrangement both the source and the test specimen were at finite distances from the lens. Taking account of this, a calculation of the value of the spherical aberration constant k_1 for a point at the extreme edge of the image yields a theoretical curve for D_c as a function of r_0 . It can be seen that the agreement between the experimental and computed curves is exact to within the experimental error. This justifies the use, in practice, of the expression $f_1 = f_{10}(1 - k_1 r_1^2)$ for the focal length of the lens. The value of k_1 obtained was 42.5 cm^{-2} which, as expected, was much higher than that for a normal projection lens ($6\text{--}8 \text{ cm}^{-2}$). Even with this high value for k_1 , ϵ^4 in (19) has a maximum value of only 10^{-4} . The value of k_1 was found to be independent of lens strength except at high values of the lens current. This gives further experimental justification of the use of this coefficient in electron optics.

Using the system with both lenses activated the distortion patterns for the same series of lens excitations were obtained. The pattern showing maximum correction of distortion is given in Fig. 12b. In the right-hand side of Fig. 14 experimental values of D_c are plotted against r_0 for four values of excitation near the point of maximum correction. Four calculated curves using $k_1 = k_2 = 42.5 \text{ cm}^{-2}$ and the indicated values of $f_{10} = f_{20}$ are plotted in the left-half of Fig. 14. The correspondence is seen to be quite good. From the geometrical arrangement of the apparatus and from the magnification at the center of the image it is possible to calculate the values of the focal lengths of the lenses if they are assumed to be equal. However, on carrying out the calculations it was found that the values of f_{10} obtained did not agree with those expected from the calculated curves. The discrepancy which was greater than the experimental error was probably partly caused by the fact that the leakage fields in the coil system used destroyed

the equality of the focal lengths of the two lenses and partly by the fact that the lenses used were not strictly thin lenses.

It should be pointed out in passing that some rotational distortion is present in the pattern of Fig. 12, but that it is of negligible magnitude.

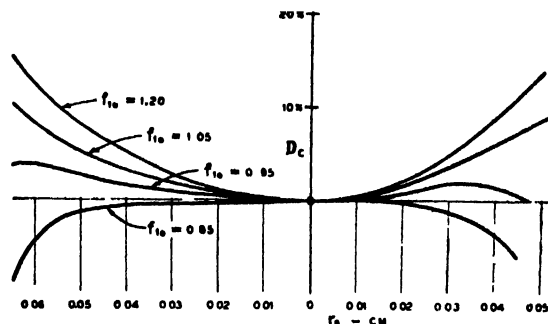


FIG. 14. Comparison between calculated and experimental values of the circumferential distortion for the double lens polepiece. In the left half, calculated values for the designated focal lengths of the lenses are given. In the right-hand half a corresponding series of experimental curves is given.

CONCLUSIONS

It has been shown in the above discussion that relatively small amounts of distortion in the projection lenses of electron microscopes can produce serious disturbances in particle size distributions obtained with the instrument. Furthermore, it has been shown that it is both theoretically and practically possible to design a projection lens system for an electron microscope in which, at least at one magnification, the distortion due to spherical aberration of that lens is reduced to negligible proportions.

ACKNOWLEDGMENTS

The author gratefully acknowledges his indebtedness to R. F. Baker and J. M. Morgan who made many of the earlier and more tedious measurements of distortion. He also wishes to thank G. W. Brown for many helpful discussions.

Parafocusing Properties of Microcrystalline Powder Layers in X-Ray Diffraction Applied to the Design of X-Ray Goniometers

J. C. M. BRENTANO

Northwestern University, Evanston, Illinois

(Received December 17, 1945; Revised February 25, 1946)

A distinction is drawn between the focusing properties of single crystals with regard to x-ray diffraction and the parafocusing properties of microcrystalline powder layers. A discussion of these parafocusing properties and of the absorption term, which arises in the diffraction of x-rays from powder layers, is followed by a discussion of various types of powder goniometers. Their characteristic properties are compared. The application of parafocusing conditions to the design of a goniometer for quantitative evaluations and of back reflection cameras is discussed in greater detail.

SOME recent developments of structure analysis and the extension of the identification method in the direction of quantitative evaluations have stressed the demand for the quantitative measurement of the relative intensities of x-ray reflections from microcrystalline powders and have given greater significance to the so-called focusing methods in which the effect of absorption on relative intensities can be eliminated. The larger angular apertures used in these arrangements as compared with the original powder rod camera is also of particular interest in industrial applications where the reduction of exposure times is relevant. While the Debye-Scherrer-Hull method, which has the inherent merits of simplicity and of utilizing minute quantities of material has been largely standardized, the requirements mentioned above have given rise to a variety of methods. It is proposed to discuss in the present paper¹ some basic points which arise with these methods with the object of bringing out their characteristics and possibly furthering the development of arrangements best suited to meet specific conditions.

1. X-RAY DIFFRACTION FROM POWDERS

The term focusing is sometimes applied to two distinct conditions of x-ray diffraction. Without wishing to break with the widely accepted use, it may be helpful for the present discussion to set out the distinction between real focusing and the ray-collective properties of powder arrangements which more appropriately

can be described as parafocusing. The distinction has not only bearing on the underlying geometrical-optical relations, with which we are here mainly concerned, but also on the purpose for which one and the other type of focusing can effectively be used.

The term focusing as defined in geometrical optics is related to the existence of an object and image space in such correlation that to a ray in the object space corresponds a ray in the image space and that to a point in the object space, defined by the intersection of a bundle of rays, is correlated a point in the image space in which the corresponding rays intersect. In x-ray optics of single crystals such a correlation can be established for monochromatic beams of wavelength λ on the ground of a "reflection" from a set of lattice planes of spacing d satisfying the Bragg law:

$$\sin \theta = n\lambda/2d. \quad (1)$$

Examples for this are the single bent crystal, the Cauchois monochromator, and the Bragg single crystal goniometer, all having focusing properties though limited to beams in one plane. No such image to object relation applies to the ray-collective arrangements of crystal powders, which may be described as parafocusing. Parafocusing is concerned with the geometrical locus of the powder particles and by disregarding their orientation establishes a necessary, not a sufficient, condition. A closer reference to the Bragg single crystal goniometer and to the parafocusing property of a flat powder layer may be helpful in bringing out the difference. In the Bragg single crystal goniometer the face of an

¹ Paper given at the Ann Arbor Summer Meeting of the American Society for X-Ray and Electron Diffraction, 1943. Abstract in *Phys. Rev.* **64**, 312 (1943).

extended crystal is so placed as to comprise the axis of rotation C (Fig. 1) of the crystal, when two points A and B in a plane normal to the axis at equal distance from C with the angle $ACB = \pi - 2\theta$ are so correlated that when the crystal is rotated various points of its surface are brought in succession in such positions as to reflect the radiation incident in the plane ACB according to (1) from A to B . The whole crystal surface does not reflect simultaneously, but true focusing is satisfied in the sense that whenever the particular reflection defined by (1) occurs, the reflected ray will go through the same point B . The flat crystal face which comprises the axis C satisfies this condition exactly. Comparing this with the conditions which arise if we substitute for the single crystal a flat layer of randomly oriented microcrystalline particles, we find that the relation peculiar to the Bragg single crystal goniometer, which limits the reflection to the particular crystal setting for which it will be directed to B , does not apply, but that for each volume element of the powder the reflection (1) will take place according to the surface of a cone of semi-aperture 2θ , having the incident beam as axis. Of this diffracted radiation some will be received at B in so far as a generatrix of the diffraction cone exists which goes through B . The fraction of the radiation so diffracted will always be very small. Various volume elements will give additive contributions to the reflection (1) received at B if generatrices of the diffraction cones for the various elements intersect in B ; for the points of a flat surface this is satisfied only in approximation, the correct surface for the powder being a surface of double curvature. This additive effect then takes place simultaneously for all parts of the powder and the best superposition is obtained for one particular setting of the layer. While the relation for the true focusing from a single crystal can only be satisfied when the distances of A and B from C are equal, the approximate para-focusing condition established by the writer for a flat powder layer, relation (3) of Section 6, applies equally well for any ratio of distances AC and BC . The intensity relations are also entirely different. For true focusing from the single crystal the intensity of the reflected radiation is independent of the crystal size and angular aperture of the

incident beam so long as the angular range, through which the crystal is rocked, does not exceed the range within which some part of it is reflecting; for the crystal powder for which all parts of the layer contribute simultaneously to the reflection, the intensity increases with the aperture of the incident beam and the area of the crystal surface. The geometry of the reflection from the single crystal rotated through an extended angular range provides that, when true focusing is satisfied, the reflections for different angles 2θ resulting for different λ 's or different d 's are recorded on a photographic film on a cylindrical surface along AB as sharp marks; while with the crystal powder, in recording an extended angular range, blurring occurs whether the powder is kept stationary or whether it is rotated unless a special screening diaphragm is used.²

In the true focusing from a single crystal, a large fraction of the reflected radiation can be directed into one particular direction or into one particular point; it is used in single crystal spectrometers and monochromators. In the para-focusing of crystal powders the reflection from the individual powder element is used at very low efficiency. It satisfies the purpose of increasing the intensities of reflections by adding the contributions from a large volume; it further makes it possible to establish powder layers which satisfy very simple absorption conditions and to average over large powder volumes. This is of significance in the quantitative evaluation of intensities and in the determination of the contour of diffraction lines.

2. PARAFOCUSING TOROIDAL SURFACE

The surface, which represents the geometrical locus of all points in space so situated that for a point source in A (Fig. 2) a ray deflected through a given angle 2θ can go through a point B , is obtained by the rotation of an arc with the

² Sir William Bragg (Proc. Phys. Soc. **33**, 222 (1921)) was able to observe the reflections from a powder layer with an ionization goniometer without taking account of the difference between the two kinds of focusing when the entrance slit to the ionization chamber acted as a screening diaphragm. Experiments of photographic recording, done in his laboratory with fatty acids, were successful (A. Muller, J. Chem. Soc. **123**, 2043 (1923)) because these presented preferential orientation and were equivalent to single crystals, but failed to give well-defined lines for a randomly oriented powder.

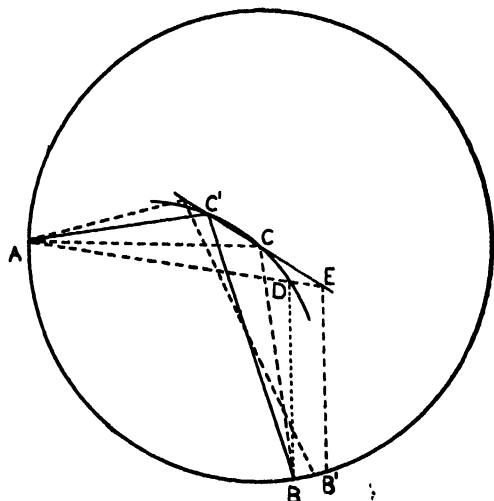


FIG. 1. The straight line through point C and C' indicates a particular setting of the face of a single crystal or of the surface of a flat powder layer going through the axis of the camera. The full line $AC'B$ indicates the reflection from the single crystal for that particular setting; as the crystal is rotated reflection takes place from various points in succession and according to Bragg's focusing relation is always directed to the same point in B . With the powder the same reflection for the fixed position occurs simultaneously from various points of the surface as indicated by the dotted lines without the Bragg focusing condition being satisfied. The curved line through $C'C$ and D marks the intersection of the plane ACB with the para-focusing toroidal surface. The projection BB' of the distance DE measures the aberration for the particular ray as discussed in Section 2.

peripheral angle $ACB = \pi - 2\theta$ about the chord AB . This surface does not satisfy the condition of focusing, but only of para-focusing: it does not establish the orientation but only the location of the crystal elements which, for a given spacing and wave-length, may contribute to the deflection of radiation from the source A to B on the ground of relation (1). This surface has been discussed by the writer in an earlier paper,³ where it was pointed out that the orientation of the lattice planes, which actually contribute to the diffracted radiation in B , is such that their normal bisects angle ACB so that they are inclined to the toroidal surface, except for the median annular zone in C' . This is the orientation given to the reflecting lattice planes in the bent crystal monochromator. The angle 4θ between the tangents to the surface in A determines the largest possible aperture of a beam from A which can be collected in B by one deflection through

³ J. C. M. Brentano, Arch. Sci. Phys. et Nat. [4] 44, 66 (1917).

the angle 2θ . With single crystals only equatorial rings or longitudinal bands tangential to the para-focusing surface can be made while with a powder it is not practical to utilize the full extent of this surface in an instrument intended to cover a range of deflection angles. Reference to this para-focusing surface, however, is useful in evaluating the efficiency of powder arrangements which approximate parts of the para-focusing toroid and offers a convenient way for determining the aberrations which result with the actual surface used. The aberration for any particular point of the powder layer is obtained by projecting its distance from the corresponding point of the toroidal surface on the plane in which the diffraction pattern is recorded. This is indicated by Fig. 1 in which DE is the distance between corresponding points of the toroid and powder element for the incident ray ADE and BB' measures the aberration. This way of evaluating aberrations saves considerable time compared with direct calculation.

The contribution in B resulting from any particular part of the toroid when represented by the surface of a powder layer is further determined by an absorption term in the intensity expression. For a narrow beam directed to a surface element of a powder layer with the absorption coefficient μ , which is sufficiently thick to be all absorbing, the absorption term is⁴

$$\frac{1}{\mu(1 + \sin \alpha / \sin \beta)}, \quad (2)$$

where α and β are the glancing angles of incidence and of reflection.

⁴ J. C. M. Brentano, Phil. Mag. 4, 620 (1927); 6, 178 (1920); Proc. Phys. Soc. 47, 932 (1935). For an all-absorbing layer the full expression for the diffracted intensity P_b received at the distance b by the entrance slit of an ionization chamber of unit height and such width as to comprise the whole reflection is

$$P_b = \frac{I p_{(hkl)} F_{(hkl)}^2 N^2 q_{(hkl)} f}{8\pi b \sin \theta \mu (1 + \sin \alpha / \sin \beta)},$$

while the corresponding expression for a small volume element dV with negligible absorption is

$$dP_b = \frac{I p_{(hkl)} F_{(hkl)}^2 N^2 q_{(hkl)} dV}{8\pi b \sin \theta}.$$

In these expressions I is the intensity of the incident beam per unit cross section and f is its cross section at the powder. The other symbols have the usual meaning as used in the last paper quoted. In this paper a misprint has actually occurred in the expression equivalent to P_b ((5) in the paper); the expressions related to it are given correctly in the paper.

3. REQUIREMENTS WHICH ARISE IN THE EVALUATION OF X-RAY POWDER DIFFRACTION PATTERNS

For *crystal structure analysis* single crystals present the advantage of giving directly the correlation between the individual reflections and the orientation of the reflecting lattice planes in the crystal. Powder diffraction methods are resorted to, apart from such instances where large single crystals are not available, when it is desired to obtain relative intensity data free from extinction from powder particles for which the size and regularity of the microcrystalline domains is so reduced as to make primary extinction small. Powder methods in the form of the quantitative powder admixture⁴ provide a simple way for interrelating the intensities of two constituents of a mixed powder to the scattering factors and to the ratio of the masses of the constituents and for obtaining absolute intensities of reflections by the admixture of a standardizing substance, for which the unit cell scattering factor is known. The intensities of the reflections to be studied can then be related to those of the standardizing admixture; they can thus be expressed in absolute measure without the necessity of measuring the intensity of the incident beam and to correct for absorption.

Crystal structure analysis requires (a) that one single record should cover an extended angular

range, (b) that the deflection angles should be recorded to a high degree of accuracy and preferably that line distances should be proportionate to angular distances so as to facilitate interpolations, and (c) that relative integrated intensities of reflections for different angles should not be affected by the shape and absorption of the particular specimen. It is further

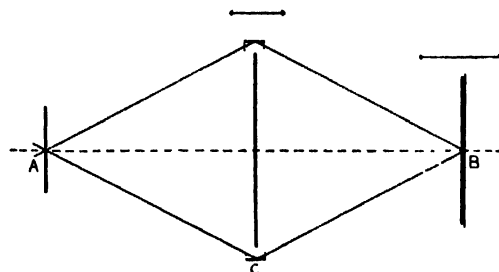


FIG. 3. Camera using an annular powder layer. For a fixed position of the ring each point of the axis is associated with a definite deflection angle.

sometimes required (d) to obtain diffraction patterns with very small quantities of material, while in another group of measurements it is desired (e) to make use of large powder volumes. This last requirement arises when the powder contains large crystalline domains of say more than 10^{-5} cm diameter when it is desired to suppress the "spottiness" of lines which arises under such conditions when the diffraction pattern is obtained from a small powder volume.

For *evaluating the degree of regularity and the size distribution of microcrystalline domains* from the line width, or better from the line contour, it is essential that (f) the line contour should not be affected by the shape and absorption characteristics of the specimen or that it should be possible to allow for such effects. It is then of particular significance to suppress irregularities of intensity distribution produced by the presence of large crystallites as indicated under (e). This implies the use of large powder volumes, incident beams of large aperture and rotation of the specimen.

For *the method of identification*, in which chemical constituents or phases are identified by their characteristic x-ray diffraction patterns in the form developed by Hanawalt, Rinne, and Frevel and by Davey,⁵ the requirements are

⁵ J. D. Hanawalt, H. W. Rinne, and L. K. Frevel, *Ind. Eng. Chem. Anal. Ed.* **10**, 457 (1938) and W. P. Davey, *J. App. Phys.* **10**, 820 (1939).

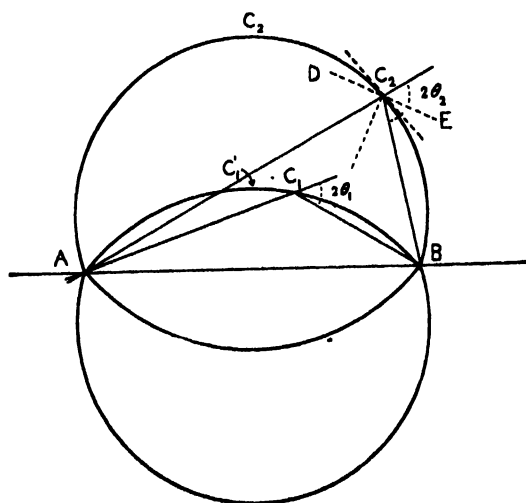


FIG. 2. Lines $AC_1'C_1B$ and $AC_2'C_2B$ indicate the intersections of two para-focusing toroidal surfaces for the angles $2\theta_1$ and $2\theta_2$ with the plane of drawing. The dotted line DE indicates the orientation of the diffracting lattice planes for the point C_2 , the normal of which bisects angle AC_2B .

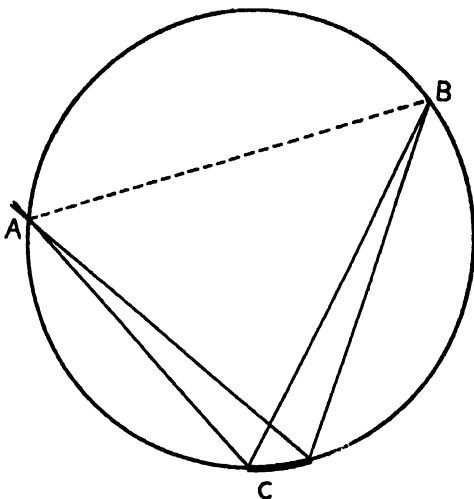


FIG. 4. Bohlin-Seemann type camera. The powder in *C* forms part of a cylindrical surface which approximates toroids constructed by rotating arc *ACB* about cord *AB* where *A* is the source and *B* for different θ 's is located at various points of the circle.

similar to those of structure analysis, only to a less exacting degree.

In *technical applications*, whether they comprise the recording of line contours or of identification patterns, the speed of recording, which is only a minor factor in structure analysis, assumes major significance. With the Debye-Hull camera and photographic recording the exposure times are longer than the times required for spectroscopic analyses or for simple chemical tests. In the field of identifications x-ray methods are thus limited to instances where the more rapid methods are inadequate, e.g., the distinction between different morphological phases or between constituents which cannot easily be separated when present in a mixture.

By using counters or ionization chambers instead of photographic recording the sensitivity can be considerably increased. It should be noticed, however, that applied to the recording of weak reflections the process of point to point scanning used with these methods necessitates slowing the scanning rate sufficiently to secure that the count numbers with which the intensities for the individual angular elements are recorded are great enough to suppress large statistical fluctuations. This limits the speed at which such measurements can be taken, particularly when the intensities are small. Industrial applications using photographic, counter, or

ionization recording set thus the demand for utilizing beams of high intensity and large angular aperture so as to reduce the exposure times.

4. THE DEBYE-SCHERRER-HULL CAMERA

The demands set out in the previous section may be considered in relation to the classical Debye-Scherrer-Hull camera. This gives patterns covering an extended angular range with proportionality between angles and line distances. The line width so far as resulting from collimating conditions is of the order of the diameter of the x-ray beam incident on the powder specimen. Very sharp lines can thus be obtained at the expense of exposure time by the use of collimating pinholes of small diameter. The effect of absorption introduces some complication, both in the evaluation of spacings and of intensities. A discussion of these conditions has been given in a paper by Möller and Reis⁶ and specific methods have been developed by McKeehan,⁷ Claassen,⁸ Bradley and Jay,⁹ and others. Recently a detailed analysis of such corrections has been given by Warren.¹⁰ Powder rods which simplify the allowance for absorption have been proposed by Greenwood¹¹ and by Rusterholz.¹² The limi-

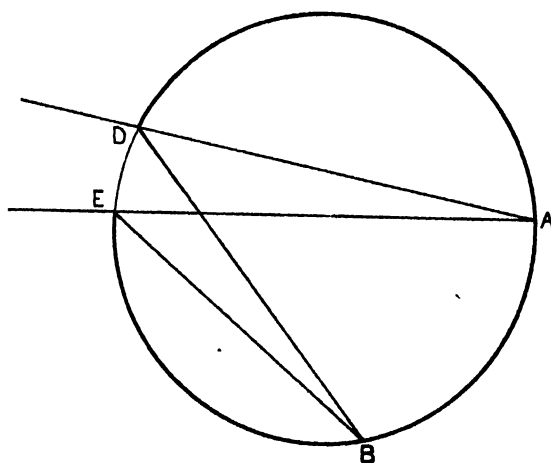


FIG. 5. Guinier type camera. The incident beam passes through the powder layer *DE* and converges toward point *A*. The diffracted beams are collected at points *B* of a cylindrical surface.

⁶ H. Möller and A. Reis, *Zeits. f. physik. Chemie* **A139**, 425 (1928).

⁷ L. W. McKeehan, *J. Frank. Inst.* **193**, 231 (1922).

⁸ A. Claassen, *Phil. Mag.* [7] **9**, 57 (1930).

⁹ A. J. Bradley and A. H. Jay, *Proc. Phys. Soc.* **44**, 563 (1932). A. J. Bradley, *Proc. Phys. Soc.* **47**, 879 (1935).

¹⁰ B. E. Warren, *J. App. Phys.* **16**, 614 (1945).

¹¹ G. Greenwood, *Phil. Mag.* [7] **3**, 963 (1927).

¹² A. Rusterholz, *Helv. Phys. Acta* **4**, 68 (1931).

tation of these procedures is that they depend on making use of a powder rod of definite shape and definite absorption, which is difficult to obtain. The effect of absorption in modifying the relative intensities can be made small by using a powder rod or wedge of negligible absorption, either by reducing its thickness or by diluting it and by the use of penetrating radiation. It will be seen that under such conditions, which reduce absorption effects, the exposure times become excessively long. In practice a compromise is made between the elimination of absorption errors and the permissible exposure times. With very thin or diluted specimens the effect of large crystallites in producing irregular lines is accentuated. It can be reduced by displacing the powder rod longitudinally or rotating it transversally.¹³

The Debye-Hull camera requires a minute quantity of material and permits to correct for faulty centering of the specimen.¹⁴ Detailed descriptions of cameras have been given by Bradley, Lipson, and Patch¹⁵ and recently by Buerger.¹⁶ Several commercial types are also available.

It will be seen that this camera can satisfy most of the demands of structure analysis and of line contour measurements and this should be considered in comparing it with other arrangements described further on. Its limitations are the long exposure times and the difficulty of making correct allowance for absorption in determining relative intensities taken over a wide angular range. When using small or diluted specimens to reduce the absorption effect the line contours are sensitive to the presence of large crystallites.

5. CAMERAS USING CYLINDRICAL POWDER STRIPS

Approximations to the toroidal surface adapted to a range of deflection angles can be obtained with rigid cylindrical rings or bands.

¹³ C. S. Barrett and A. G. Guy, *Rev. Sci. Inst.* **15**, 13 (1944). E. Schwarz, *Phys. Zeits. Sowietunion* [5] **3**, 443 (1934).

¹⁴ M. Straumanis and O. Mallis, *Zeits. f. Physik* **94**, 184 (1935) and M. Straumanis and A. Jevins, *Zeits. f. Physik* **98**, 461 (1936).

¹⁵ Bradley, Lipson, and Patch, *J. Sci. Inst.* **18**, 216 (1941).

¹⁶ M. J. Buerger, *J. App. Phys.* **16**, 501 (1945). See also *X-Ray Crystallography* (John Wiley and Sons, Inc., New York, 1942).

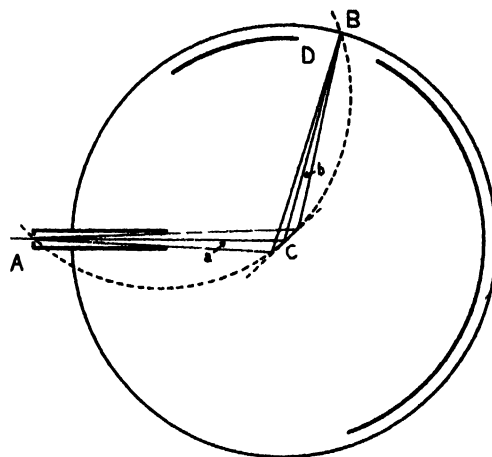


FIG. 6. Parafocusing central powder layer.

As a derivation from the toroidal surface the writer proposed an annular band.¹⁷ The powder is placed on the inside surface *C* of a ring (Fig. 3) the axis of which comprises the points *A* and *B*. The radiation from *A* will then be deflected to *B* for a range of angles which can be varied by displacing the ring parallel to itself and displacing point *B* along the axis. *B* is represented by a small opening in a screen closely behind which the photographic film is placed. With the displacement of the ring *C* and of *B* the photographic film is moved at right angles to the axis *AB*. In praxis only part of the ring is used. The arrangement gives very high intensities. It does not, however, comply well with other demands when photographic recording is used because of the oblique incidence of the diffracted x-rays on the film, which is unfavorable for quantitative measurements. The arrangement may be useful in combination with a counter or ionization chamber.

A longitudinal strip is used in the arrangement of Seemann¹⁸ and Bohlin¹⁹ (Fig. 4) in which the toroidal surface is approximated by a cylindrical area tangential to a longitudinal band of the toroid. This arrangement is very simple in design and capable of recording a large angular range on one single film strip. Line distances are not proportionate to angular distances; it is unfavorable for intensity measurements and the meas-

¹⁷ J. C. M. Brentano, *Arch. Sci. Phys. et Nat.* [5] **1**, 550 (1919).

¹⁸ H. Seemann, *Ann. d. Physik* **59**, 455 (1919).

¹⁹ H. Bohlin, *Ann. d. Physik* **61**, 421 (1920).

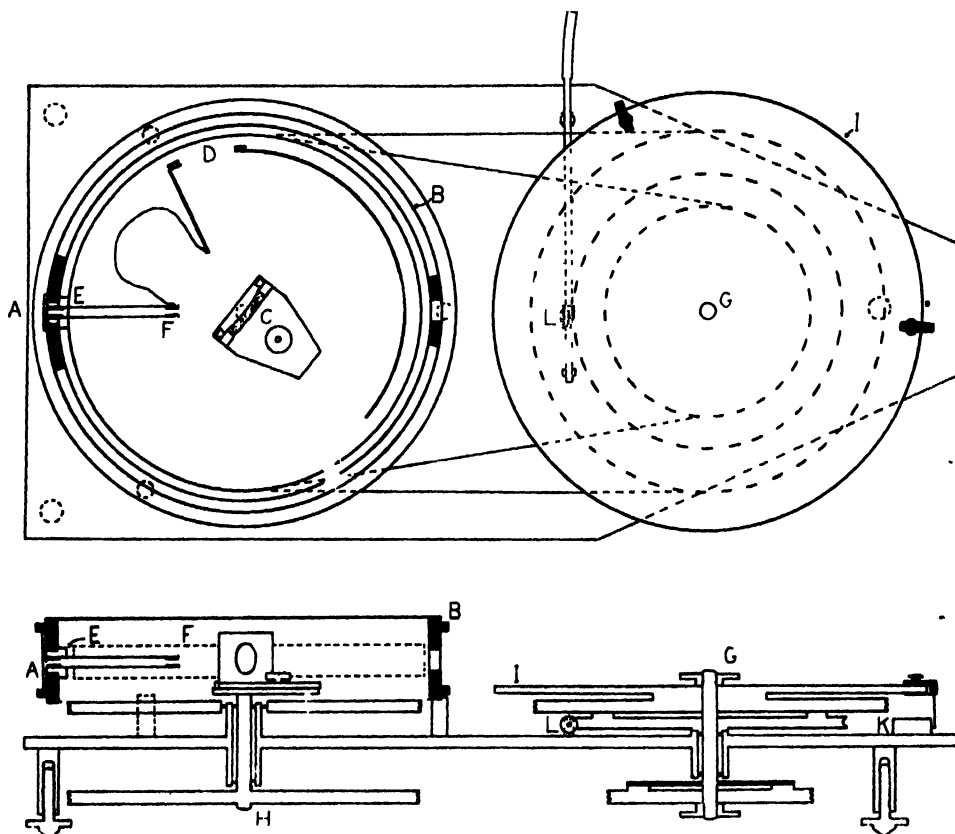


FIG. 7. Schematic diagram of camera with central parafoocusing powder layer. The film is wrapped round the cylindrical camera body *B* and is exposed through slots. The collimating system consists of a slit in a front plate *A* and of inserts *E* and *F* in the collimating tube. Insert *E* determines the slit height; insert *F* determines the angular aperture. Slit and inserts are exchangeable. Aperture *D* can also be varied. A reversible motor through worm drive in *L* rotates shaft *G*. Other details are given in reference 24.

urement of line contours because the diffracted rays strike the photographic film with varying obliquity and because the entrance diaphragm in *A* is reproduced from the various parts of the powder surface at different rates of magnification. The method is, however, well adapted for utilizing large volumes of powder in conjunction with incident beams of large aperture. Several asymmetrical cameras of this type have been described. De Jong²⁰ described a symmetrical camera, the cylindrical powder surface being opposite to the entrance slit. This is particularly suitable for back reflection records.

An arrangement, in which the photographic film forms a cylindrical surface while a thin powder layer forms a small area of the same surface (Fig. 5) and a converging incident beam

is transmitted through the powder with the center of convergence situated on the cylindrical surface, has been proposed by Guinier²¹ who makes use of a monochromator for obtaining the converging beam. Frevel²² has proposed a similar arrangement using a converging Soller slit. While this arrangement can be discussed in terms of parafoocusing surfaces, the most direct treatment results from considering that for the plane of incidence the condition of equal convergence of the incident and diffracted beams is satisfied for points of convergence situated on the cylindrical surface. The characteristics of this method are similar to those of the Bohlin-Seemann camera. Its particular merit is that very small diffraction angles can be observed while, on the other hand,

²⁰ P. de Jong, *Physica* 1, 23 (1927).

²¹ A. Guinier, *Comptes rendus* 204, 1115 (1937).

²² L. K. Frevel, *Rev. Sci. Inst.* 8, 475 (1937).

the incident beam is transmitted through the powder so that the intensities depend on the absorption of the powder. The arrangement of Bohlin-Seemann, de Jong, and Guinier satisfies the parafofocusing condition essentially with respect to one plane.

6. GONIOMETER WITH A CENTRAL POWDER LAYER

A goniometer proposed by the writer²³ uses a powder layer in the center of a camera which satisfies the parafofocusing condition for points of a cylindrical surface. It is intended to conform with those requirements set out in Section 3 which are satisfied in the Debye-Hull camera and to provide in particular proportionality between line distances and angles and normal incidence of the diffracted x-ray beam on the photographic film. In addition it establishes such conditions that the relative intensities of the reflections from thick powder layers are independent of the particular absorption characteristics of the powder specimen and are actually the same as would be obtained from a small powder volume with negligible absorption. It uses larger powder volumes and wider angular apertures than the Debye-Hull camera and requires shorter exposure times although they are longer than with the arrangements of Section 5. A detailed discussion of the geometrical-optical relations and aberrations resulting with this method has been given in an earlier paper²⁴; it will here be sufficient to indicate its characteristic features.

Considering in the first place the case of photographic recording, it will be seen (Fig. 6) that the toroidal surface intersects the surface of the photographic film. The point of intersection *B* determines the angle for which the parafofocusing condition is satisfied for a powder layer in *C* so oriented as to conform with the particular parafofocusing toroidal surface. The recording of an extended angular range requires then a screen with an opening *D* which in succession uncovers different parts of the film surface as the powder layer is rotated about an axis normal to the plane *ACB*. If *a* and *b* are the distances of the powder layer from the collimating slit *A* and the

cylindrical surface *B* and α and β are the glancing angles of the incident and reflected beam with the plane tangential to the powder surface in *C*, the parafofocusing condition is satisfied when

$$\sin \alpha / \sin \beta = a / b. \quad (3)$$

The displacements of the powder layer and of the screen opening *D* are so correlated as to conform with the constant $\sin \alpha / \sin \beta$ ratio of (3) for the whole angular range. It will be seen that this constant ratio introduced in the absorption term (2) secures that this term has the same values for all angles and that relative intensities are independent of the absorption characteristics of the powder.

The parafofocusing powder surface is in a general way a surface of double curvature, which changes with the diffraction angle. A first approximation to it is a flat surface; a much better approximation adapted to a particular range is a surface of double curvature—data for radii of curvature of such surfaces are given in the second paper²⁴—and a spherical surface, when an extended angular range is to be covered and the angular aperture of the incident beam is not too large. For $a=b$ for the 2θ range from 60° to 150° a spherical surface with a radius of curvature $3/4b$ will give considerably better definition than a flat layer, permitting both in the horizontal and in the vertical plane²⁵ an incident beam of an angular aperture of 2° – 3° . This range can be extended by a V-shaped horizontal slit in front of the collimator near to the powder which moves with the powder layer. Such a slit can be used to restrict the vertical aperture for small angles; it is also helpful in reducing the strong intensities in that range. When using such a slit, allowance for it must be made in computing intensities.²⁶ In a more elaborate arrangement the vertical and horizontal aperture of the beam can be so controlled as to keep the solid angle of the incident beam constant and to restrict either the horizontal or vertical aperture in such a way as to obtain the smallest aberration. Such devices

²³ For brevity in the following discussion, the term *horizontal* is used for the plane *ACB* in which the reflections are observed and the term *vertical* for the plane which comprises the axis of rotation of the powder layer.

²⁴ For accurate intensity measurements we prefer in general to use a beam of constant vertical aperture for the whole range.

²³ J. C. M. Brentano, Proc. Phys. Soc. 37, 184 (1925).

²⁴ J. C. M. Brentano, Proc. Phys. Soc. 49, 61 (1937).

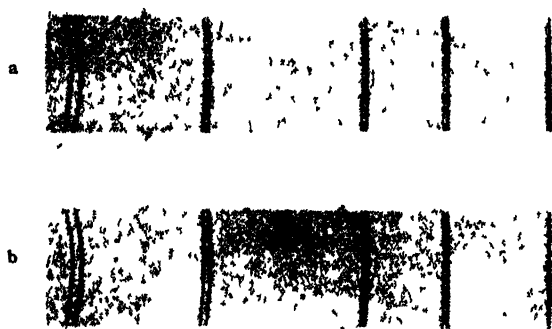


FIG. 8. Group of cadmium oxide lines recorded with parafofocusing central powder layer (a) with rotating diaphragm, (b) without rotating diaphragm. Spherical layer: angular aperture of incident beam 2° .

are, however, of interest only then when it is desired to use beams of considerable angular aperture.

The schematic diagram (Fig. 7) indicates one design of the goniometer. The powder layer in C and a platform, on which the diaphragm D is mounted, rotate about a common axis II . They are driven by means of pulleys and steel wires from a shaft G . In combination with good ball bearings this system is very satisfactory. Projecting stops on a graduated disk I act on a microswitch K and determine the range of motion. For the ratio $a:b=1$, i.e., when the collimating slit A is on the cylindrical surface of the camera, the correct ratio of the angular displacements is 1:2. This is the simplest arrangement which can be satisfied with centric circular pulleys. The relative angular displacements corresponding to a ratio $a:b$ different from 1 can be obtained with pulleys of suitable contour. A spring loaded idler pulley is then required to maintain uniform tension. For correct intensities uniform velocity is given to the rotating screen while, when a and b differ, the angular speed of the powder layer varies. A ratio $a>b$ is used for obtaining very large apertures for the incident beam; slit A can then be moved into immediate vicinity of the target of the x-ray tube. This is discussed in Section 8.

The use of the rotating diaphragm in D implies that the exposure times increase with the angular range recorded. To compare the exposure times resulting from these conditions in recording a

given range of say 90° with those of a Debye camera of equal size, it may be assumed that the outer pinhole of the latter and the entrance slit A of our camera are of equal area and that the aperture for the parafofocusing beam is 2° by 2° , while the pinhole collimator with a length of $3''$ has a pinhole diameter of $0.020''$, as commonly used in some commercial instruments. The exposure required with the Debye camera is then 4.8 times greater than with the parafofocusing central powder layer. This takes account of a diaphragm opening of 22° in D when the motion of diaphragm D exceeds by 22° on either side the 90° range, so as to secure uniform exposure for that range. When recording a group of lines comprised within a limited angular range, the rotating diaphragm can be dispensed with. The powder layer is then so set as to satisfy the parafofocusing condition for the middle of this range. It can be rocked through a few degrees to avoid spotty lines. For the conditions considered above the intensity is then 29 times larger with the parafofocusing layer than with the Debye camera. When a moderate amount of diffusion is permitted a range up to 60° can be recorded without using a moving diaphragm; this can be increased by reducing the angular aperture of the incident beam. Figure 8 shows a group of cadmium oxide lines recorded with and without diaphragm. The angular range covered is 65° .

When recording the diffraction pattern with a counter or ionization chamber its entrance slit is situated on circle B . The slit acts as diaphragm D and eliminates those aberrations which arise for angles away from the one for which the parafofocusing relation is satisfied.²⁷ It does not eliminate the aberrations resulting from any deviation of the shape of the powder surface from the parafofocusing toroid. The line broadening resulting from aberrations is then essentially determined by the deviations of shape and orientation of the actual surface from the toroid so that the use of correct surfaces and suitable diaphragms becomes important. When using ionization chambers and counters, which must be placed behind the circle B , the use of $a>b$ makes it possible to extend the angular range recorded without interference from the x-ray tube.

²⁷ Care must be used to avoid errors in evaluating intensities which result with too narrow slits.

We find it most convenient to carry out identifications on the ground of patterns covering a 2θ range extending from 60° to 150° and to disregard the ranges of very small and very large angles. For this middle range, particularly above 90° , the lines are well defined and any broadening can best be examined here. For the range of large and small angles the curvature of the lines, which results from the small aperture of the diffraction cones and which arises with all powder methods, is unfavorable for accurate measurements; besides the number of lines in these ranges is generally small

For evaluating intensities the simplification with this arrangement results from the fact that by satisfying (3) for a fixed ratio a/b the absorption term (2) becomes independent of the diffraction angle as indicated above. The relative intensities are thus independent of the absorption coefficient of the material and are those for a small volume element with negligible absorption. They are thus directly comparable with those obtained from a highly diluted Debye powder rod. It should be noted that this independence from the deflection angles does hold good only when satisfying the parafocusing condition. When recording an extended angular range without the revolving diaphragm D the absorption becomes the function of the angular setting; it can be calculated for each line.⁴

The large angular apertures and the use of thick all-absorbing layers increase the powder volumes utilized and tend to suppress the spottiness caused by the presence of individual large crystallites. This can be further reduced by using a flat powder layer in form of a rotating disk with the axis of rotation off-set with respect to the incident beam. This disk takes the place of the powder layer in the center of the camera. The volume of powder and the angular range through which it is exposed to the incident x-ray beam is thus increased. Figure 9 shows the patterns of the same material taken with a rotated and displaced powder rod and with a rotated flat disk in the way described.

7. BACK REFLECTION CAMERAS

One particular type of goniometer is that for taking "back reflection" patterns, limited to

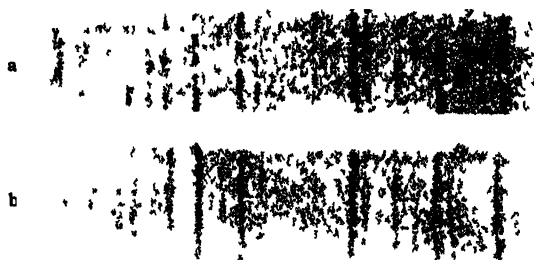


FIG 9 Patterns of the same material taken (a) with a rotated and longitudinally displaced rod and (b) with the camera of Section 6 using a rotating disk.

large 2θ values. Changes of line position and of line width and contour in that range are most sensitive to changes of lattice spacings and to imperfections of the microcrystalline domains. In a widely used type of camera²⁸ the pattern is recorded on a flat film, a pinhole collimating tube being placed between the x-ray source and the recording film. By using a narrow beam similar to that used with the Debye-Hull camera sharp lines can be obtained over a range of angles without having to conform with parafocusing conditions. As with the Debye camera the exposure times are then long.

By placing the outer collimating pinhole into a definite relation to the flat photographic film a beam of larger aperture can be used, good definition is then secured for one particular angle.²⁹ A flat or spherical powder layer in C (Fig. 10) is only an approximation to the parafocusing surface which is not spherical but has a radius of curvature $s/2$ in the horizontal plane and approximately s in the vertical plane, s being the distance from A to C . The angular aperture should thus not be too large. The merit of these arrangements is that with circular pinholes and diaphragms they possess azimuthal symmetry. They can thus be used for recording fiber structures and preferential orientations.

For a powder with random orientation a camera conforming with the toroid is preferable since it gives good definition for an extended angular range. By substituting for the cylindrical powder layer used by de Jong a spherical powder layer (Fig. 11) with a radius of curvature equal to

²⁸ G. Sachs and J. Weerts, *Zeits. f. Physik* 60, 481 (1930).

²⁹ R. Berthold, *Zeits. f. Physik* 15, 42 (1934).

$3b/2$, where b is the camera radius, a better approximation to the parafoocusing surface is obtained which takes account of the divergence of the beam in two planes and not in one plane only. Large apertures can be used with short exposure times except that for the 2θ range approximating 180° where the diffraction lines are very strongly

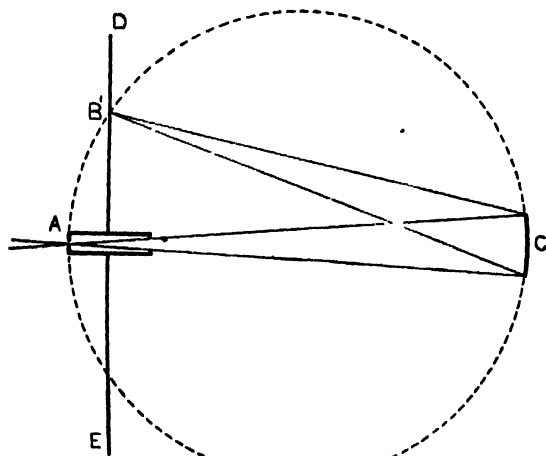


FIG. 10. Back reflection camera for recording preferential orientations. The pattern is projected on a flat photographic film in DE . The arc in C indicates the intersection with the parafoocusing surface which, however, is not spherical but a surface of double curvature. When using a flat or spherical powder surface and circular pinholes for obtaining uniform conditions for all azimuthal angles, the aperture should thus not be too large.

curved, the vertical aperture should be somewhat restricted. By rotating this spherical surface about an axis somewhat offset to the direction of the incident beam the quality of the lines, when large crystallites are present, is greatly improved. Restricting the aperture of the beam in the horizontal plane will reduce the residual aberration in that direction and extend the angular range of good definition. A slit in A can be substituted for a pinhole as previously discussed in Section 6, except that its height should be limited for recording reflections for which 2θ approximates 180° ; for such angles, however, the resolving power is so great that good definition can often be dispensed with. The line distances are not proportionate to the angles, a disadvantage which, however, is felt less for large angles since the higher resolving power makes it easier to dispense with accurate linear interpolations. The goniometer of Section 6 will record back reflection patterns and at the same

time give proportionality between angular and linear distances. When used without rotating screen with the powder layer set at right angles to the incident beam symmetrical patterns are obtained; best definition is then satisfied for $2\theta = 180^\circ$ where it is least relevant.

A more effective arrangement is obtained by placing A within the camera (Fig. 12) and setting the powder layer at such an angle as to be tangential to the parafoocusing surface for a point B' , which forms the center of the particular range for which good definition is required. It will be seen that the toroid (dotted circle) approximates the circle B on which the film is placed, so that for the projection on B a linear relationship between angles and film distances is secured, while at the same time approximating the parafoocusing condition over an appreciable range. The radius of the spherical powder surface should be approximately $b/2$. Here again for the central strip of the photographic film the residual aberration in the vertical plane is not disturbing. This arrangement is very convenient for obtaining rapid data in that angular range of 150° – 170° , which is most relevant for back reflection. While the ratio $\sin \alpha / \sin \beta$ is not constant, it varies but little with angle for large 2θ values so that the relative intensities approximate those obtained with the rotating screen.

8. COLLIMATING SYSTEMS

Three types of collimating systems are in use: The pinhole system, consisting of two pinholes placed at a distance which is large compared with their diameter, defining a narrow, approximately parallel beam. The Soller system, which gives beams with small angular divergence in one plane and large angular divergence in a plane at right angles. The system comprising a pinhole or slit which acts as a secondary source, and a second larger diaphragm which defines a beam comprising a definite solid angle.

The first is used in the Debye-Hull camera and in the usual back reflection camera, in which the location of the pinholes is irrelevant provided the collimator is sufficiently near to the powder to prevent an undue broadening of the beam. The Soller system, which in its angular spread is similar to that given by a Cauchois monochromator, meets the condition where focusing or

parafoocusing is satisfied for one plane. It gives an angular spread similar to that of a Cauchois monochromator and Guinier and Frevel have used one and the other in the arrangement of Section 5. Beams comprising definite solid angles are used when the parafoocusing condition is satisfied for two vertical planes. In comparing the merits of slits and of circular pinholes it should be considered that when the curvature of the diffraction lines is small, i.e., outside the range of very small or very large angles, some residual aberration in the direction of the lines is irrelevant. Slits oriented parallel to the direction of the lines can then be used without impairing the definition. The slit has then the advantage not only of giving greater intensity, but also of increasing the angular range through which the individual volume element is exposed to the incident x-ray beam. This improves the quality of the lines.

With any collimating system the intensity of the beam is independent of the distance of the entrance slit from the target of the x-ray tube so long as the beam, as defined by the slit, pinholes or diaphragms, projected backwards on the target covers an area which falls within the emitting anode spot. With the usual commercial tubes and pinhole diameters of 0.02" this is satisfied when the distance between the entrance pinholes and tube target does not exceed the collimator length. There is then no gain in bringing the entrance slit nearer to the target. When conforming with parafoocusing conditions with a collimator comprising a slit which acts as secondary source, the location of this slit with respect to the powder is relevant. At the same time with the much larger angular apertures used this requires bringing the slit or secondary source to the target; with tubes with small window to target distances apertures of 2° – 3° can thus be obtained. When much larger apertures are required for obtaining extremely short exposures, the slit can be placed within the x-ray tube or the target used as slit. The writer has done this with a laboratory type tube operated from a pump²⁸ which in general design is similar to the new type Machlett tube. An alternative way of satisfying the requirements of large apertures is to use an extended anode spot. With an electron

beam concentrated on a very small target area a moderate broadening of the spot does not effectively increase the loading capacity, because the volume of the anode surrounding the spot determines the rate at which heat can be dissipated; this is only little changed when the target area is increased by a small amount. With an extended spot, however, on a thin target and with forced water circulation the total loading capacity of the tube and the x-ray output can be considerably increased.

9. THE EVALUATION OF INTENSITIES

The conditions for quantitative evaluations are met most directly by the method of Section 6, which we developed primarily for the purpose of eliminating the effect of absorption on the relative intensities of the reflections recorded over an extended angular range. For identifications, so far as they depend on the relative line intensities, it is an obvious advantage if these are unaffected by the absorption in the particular powder specimen. For quantitative determinations on mixed powders²⁹ and the determination of unknown structure factors in terms of structure factors which are known, to which we have referred in the earlier part, reflections occurring for different angles can thus be correlated.

In such measurements it is essential to avoid preferential orientation of the powder particles. One way of achieving this is to place the powder with the binding material in a frame which in preparing the layer is held against a surface of the desired curvature, the powder being intro-

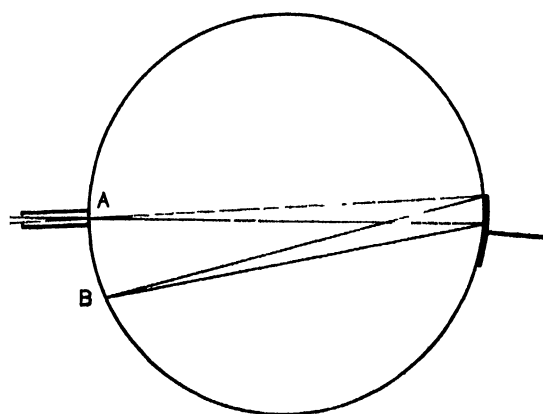


FIG. 11. Back reflection camera for recording powder diagrams on a cylindrical surface *B* using a spherical powder layer rotated about an offset axis.

duced from the back of the frame so as to avoid any sliding of the material near the front surface. It is then also required that the aperture of the incident beam, as defined by the collimator, should be fully utilized for the whole angular range and that the powder layer should be sufficiently thick to be practically all-absorbing. While the large volume of powder utilized is an advantage in improving the quality of the lines, this excludes the method in such instances where very little material is available. Eventually the condition of full absorption can be satisfied with materials constituted of light atoms by an admixture for which the absorption coefficient is high. When evaluating the relative intensities for mixed powders, the mixture should be sufficiently fine for the domains of heterogeneity to be small in terms of the distance for which appreciable absorption occurs.³⁰

The goniometer with a central powder layer can be adapted to methods relating diffracted to incident intensities³¹ and to the substitution method and to the evaluation of intensities from powder layers at fixed angles, discussed by Brindley and Spiers³² and by the writer.³³ It can also readily be adapted to the recording of the diffraction lines on a rocking film, a method recently described by the writer³⁴ which considerably improves quantitative evaluations of lines which have an unfavorable line contour.

10. THE EVALUATION OF DIFFRACTION ANGLES

One way of correlating measured line positions and diffraction angles is to measure the distances of corresponding lines on a symmetrical pattern taken on both sides of a cylindrical camera. This is generally done in the evaluation of Debye patterns. In an alternative way fiducial marks are placed on the film. In a third method reference lines are recorded on the pattern by admixing to the powder a substance for which the spacings are well known. With all these methods interpolations are facilitated when the line distances are proportionate to angular distances.

³⁰ J. C. M. Brentano, *Proc. Phys. Soc.* **47**, 932 (1935).

³¹ J. C. M. Brentano, *Zeits. f. Physik* **99**, 65 (1936).

³² Brindley and Spiers, *Proc. Leeds Phil. Lit. Soc.* **3**, 4 and 73 (1935); *Proc. Phys. Soc.* **50**, 17 (1938).

³³ J. C. M. Brentano, *Proc. Phys. Soc.* **50**, 247 (1938).

³⁴ J. C. M. Brentano, *J. Opt. Soc. Am.* **35**, 382 (1945).

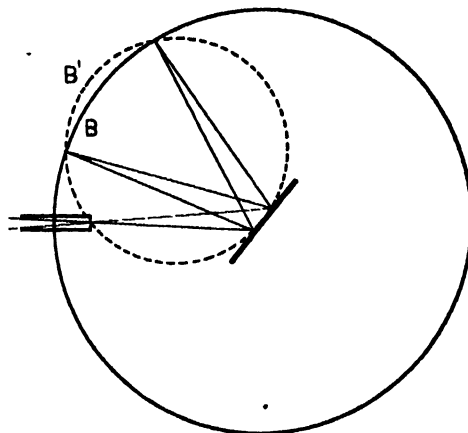


FIG. 12. Back reflection camera for recording powder photographs on a cylindrical surface B with line distances proportionate to angular distances.

By making use of symmetrical patterns in the technique of Straumanis,¹⁴ developed for the Debye-Hull camera, errors due to faulty positioning of the specimen can be eliminated. This does not eliminate residual errors which arise from the penetration and absorption of x-rays in the powder rod and errors which result from uneven film shrinkage, errors which may be neglected for ordinary purposes but which become appreciable in view of the long distances over which the comparison measurements extend when high accuracy is demanded. Fiducial marks offer the simplest evaluation in terms of points on the circumference of the camera. The location of such marks represented by edges separating dark and light fields on the film is affected by the photographic Eberhard effect, a source of errors which has been fully discussed in relation to optical spectroscopy. This error is avoided with marks consisting of V-shaped dents or projections. These can be located to the same or to a higher degree of accuracy than the diffraction lines.³⁵ Such regularly spaced marks eliminate errors due to uneven film shrinkage; they do not in themselves, however, take care of errors due to faulty positioning of the powder.

Errors of asymmetry and eccentricity should be considered as arising from two different sources. One results from faults of workmanship;

³⁵ A way of utilizing these marks in the microdensitometric evaluations is to trace with the edge of a razor blade fine scratches between the corresponding apexes of the V's.

in the camera with a parafoocusing central layer such an error would be a deviation between the axis of rotation of the powder and the cylindrical surface on which the film is placed or faulty location of the entrance slit of the counter with respect to the graduated scale on which the angles are read. By careful designing and workmanship such errors can be reduced to limits smaller than the accuracy to which line positions can be evaluated. It is thus indicated to suppress such errors to the limits of accuracy required rather than to correct for them in each exposure. The other results from faults in mounting of the individual specimen and the penetration of x-rays within the powder. Powder surfaces can be located to a fair degree of accuracy; for instance in the parafoocusing camera with central layer, the surface of powder specimens, made by the use of jigs or molds in the way described in the last section, can be located within 0.2 mm. With a camera of 6" diameter the accuracy is then sufficient for ordinary identification purposes. When a higher degree of accuracy is required, faults of location and penetration can be eliminated for this particular method by using a flat thin layer on a support (plastic or glass) transparent to x-rays²³ in combination with a beam of limited angular aperture. Two exposures are then recorded on the same film for which the powder layer has been rotated by 180°, so that opposite sides of the layer face the x-ray beam. If the axis of rotation does not go through the powder layer, double lines result, the correct position being half-way between the two peaks. The line positions defined in this way are related to the fiducial marks. The technique of using thin layers is only needed when the highest accuracy is demanded.

The admixture of a substance with known spacings eliminates all errors resulting from faulty positioning and absorption and also all other aberrations. To distinguish between the two line patterns and detect superpositions it is then often necessary to take two patterns, one with the admixture and one without.

11. CAMERA SIZES

The effect of camera sizes on the accuracy of angular measurements has been discussed in an

earlier paper,²⁴ to which reference may be made to avoid repetition. It may here be pointed out, however, that the marginal error attached to the use of finite angular apertures—which generally increases with the square of the aperture—is essentially a relative angular error; for given apertures it is thus independent of the diameter of the circle on which the pattern is projected. The same applies to the error resulting from uneven film shrinkage, which is proportional to the film length and in terms of angles independent of the camera size. Errors of this group have in common that their effect on the relative accuracy is independent of the camera size. Errors of a second group, resulting from the diffusion of x-rays in the photographic material, from the penetration in the powder, and essentially also from faulty location of the specimen, involve a certain marginal linear uncertainty in the location of points on the film surface. The effect of such errors on the angular accuracy decreases with the camera size. For best efficiency such conditions will be established as to obtain a balance in which no single error is preponderant. The statement sometimes made in regard to Debye-Hull cameras, that the exposure times increases with the third power of the camera diameter, is thus without definite significance without indicating whether the pinhole diameters and collimator lengths are changed in proportion to the camera diameter. If all dimensions are changed in proportion, the angular apertures of the beams are maintained and an increase of camera size satisfies essentially the purpose of reducing the effect of the second group of errors without much change of exposure times. If, on the other hand the same pinholes are used with a larger camera in which the length of the collimator tube has been increased in proportion to the camera diameter, some of the aberrations are very effectively decreased. Such a change leads, however, to rapidly increasing exposure times; applied to a "balanced" setting it leads to conditions in which some of the errors are unduly suppressed for best efficiency. An adjustment to the specific conditions is demanded in each case.

We generally chose the diameter best suited for our requirements and neglected to pay

attention to standard camera dimensions, since we found it not too difficult to have a scale made so adjusted to the camera diameter that line distances can be read directly in degrees. It is, however, possible to adjust slit openings and apertures so as to yield best efficiency for a given camera diameter. For microdensitometric evaluations the broader lines, which for a given

angular resolution are obtained with a larger camera, are equally good or more favorable than the sharper lines with a small camera and we prefer thus large camera diameters in order to suppress aberrations of the second group. For visual observation the greater density gradients resulting with a small camera giving a close pattern of fine lines is a distinct advantage.

A Study of Oxide Cathodes by X-Ray Diffraction Methods

Part I. Methods, Conversion Studies, and Thermal Expansion Coefficients

A. EISENSTEIN*

Radiation Laboratory,† Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received November 13, 1945)

Two methods are described for obtaining x-ray diffraction patterns of oxide-coated cathodes. One method is used in the study of the conversion process in forming the oxide cathode while the other method is utilized in measuring the thermal expansion coefficients of barium, strontium, and thorium oxides. The conversion of an equal molar barium-strontium carbonate solid solution, $(\text{BaSr})\text{CO}_3$, involves (1) crystal growth in the carbonate, (2) decomposition to the mixed oxides, BaO and SrO , (3) formation of the oxide solid solution, $(\text{BaSr})\text{O}$, and (4) crystal growth in the oxide. A similar sequence of events is observed in the conversion of a mixed carbonate, $\text{BaCO}_3 + \text{SrCO}_3$. Crystal and particle size growth of carbonates and crystal growth of oxides are investigated and possible relationships are discussed.

I. INTRODUCTION

THE thermionic emission properties of oxide cathodes are directly related to certain chemical and physical structures; it is, therefore, essential for future experimental and theoretical studies that an accurate description of these structures be available. X-ray and electron diffraction studies have supplied information regarding the chemical nature and physical structure of active oxide coatings which is obtainable by no other means.

Burgers¹ showed that when a mechanical mixture of barium carbonate and strontium carbonate was decomposed in vacuum during the formation of an oxide cathode, a simple conver-

sion to the mixed oxides, BaO and SrO , took place. Further heating caused the formation of a true solid solution, $(\text{BaSr})\text{O}$, at a rate which was temperature dependent. The lattice constant of such a solid solution was found to vary linearly with composition and continuously from 100 percent BaO to 100 percent SrO . Cathodes of $(\text{BaSr})\text{O}$ which were flashed at a high temperature, when examined by x-rays, showed asymmetrical diffraction lines which were interpreted as indicating a preferential loss of BaO from the surface.

Benjamin and Rooksby² found that a coating which was initially a carbonate solid solution, $(\text{BaSr})\text{CO}_3$, decomposed to give an oxide solid solution, $(\text{BaSr})\text{O}$. A plot of d.c. thermionic emission versus relative proportions of BaO and SrO present in the solid solution indicated a dependence such that the optimum emission was ob-

* Now at the Research Laboratory of Electronics, Massachusetts Institute of Technology.

† This paper is based on work done for the Office of Scientific Research and Development under contract OEMar-262 with the Radiation Laboratory, Massachusetts Institute of Technology.

¹ W. G. Burgers, *Zeits. f. Physik* **80**, 352 (1933).

² See e.g., J. P. Blewett, *J. App. Phys.* **10**, 668 (1939).

² M. Benjamin and H. P. Rooksby, *Phil. Mag.* **15**, 810 (1933).

served from a composition of about 60 mole percent SrO–40 mole percent BaO. A later investigation⁴ pointed out the relationship between the physical state of the oxide and the thermionic emission: higher emissions were obtained when the oxide was homogeneous throughout than when it consisted of two or more phases. Benjamin, Huck, and Jenkins⁵ used x-ray diffraction to supplement optical microscope and emission microscope data in a study of the physical state, crystal structure, and emission properties of coatings prepared under varied conditions.

One hope for the diffraction type experiment has been to check the theory⁶ of electron emission which supposed a nearly monomolecular layer of "free" alkali metal on the external surface of the semi-conductor oxide. Because of the relatively low penetrating power of electrons as compared to x-rays, electron diffraction methods are best suited to the examination of surface layers. Gaertner,⁷ using this method, confirmed the findings of Burgers: namely, that a cathode of bulk composition BaO and SrO has a surface of pure SrO, some 300 to 2000 atomic layers thick,⁸ formed immediately following the conversion from the carbonates. Darbyshire⁹ found similar surface layers on cathodes of bulk composition (BaSr)O and SrO; however, cathodes converted from BaCO₃ gave surface patterns of BaO. Diffraction patterns taken during the conversion of the carbonate solid solution showed lines of only (BaSr)CO₃ or (BaSr)O. If an adsorbed monomolecular surface layer of Ba were bound to underlying O lattice sites of SrO, it would be expected to give rise to a diffraction pattern differing from that of SrO only in relative line intensities. A comparison¹⁰ of the SrO surface patterns from cathodes of (a) (BaSr)O and (b) SrO suggests the possible presence of an active layer in the first case. A more recent critical examination⁸ of such patterns showed no evidence for the supposed active monomolecular

layer. One should regard both results with a degree of skepticism since (1) although a layer of Ba at lattice sites corresponding to the Sr of SrO has a scattering power of roughly twice that of Sr, it is only one of the some 20 surface layers contributing to the total scattering, and (2) an adsorbed layer of randomly spaced Ba would contribute only to the diffuse scattering.

Part I of this paper will discuss methods which were found useful in the further study of oxide coated cathodes by x-ray methods. The application of these techniques to the study of chemical and physical changes during conversion will be presented as well as the application to measurements of the thermal expansion coefficients of the oxides.

II. METHODS

Since the oxides of barium and strontium are chemically unstable in an atmosphere containing carbon dioxide or water vapor, special techniques were required to permit diffraction patterns of the active cathode coatings to be taken without exposing the cathode to the atmosphere. Two methods were used: the first, which consisted essentially in guarding the cathode against atmospherics by a protective coating, is particularly applicable to the study of a great many different cathodes removed from vacuum tubes; the second, which employed a special x-ray diode, allows a detailed study of single cathodes as a function of processing and life. The first method may also be used to investigate changes which occur as a function of life if a separate cathode is used for each point studied.

Although a method has been reported³ for removing activated oxide cathodes from special test diodes to thin-wall, evacuated glass capillaries for x-ray examination, this technique is not suited to the examination of large numbers of cathodes nor to cathodes processed in standard vacuum tubes. More easily adapted to this study is a method⁸ consisting of opening the vacuum tube in an airtight glass-windowed manipulation box in which the atmosphere is maintained chemically inert by an adequate flow of carefully dried nitrogen.* Rubber gloves, attached to the

⁴ M. Benjamin and H. P. Rooksby, *Phil. Mag.* **16**, 519 (1933).

⁵ M. Benjamin, R. J. Huck, and R. O. Jenkins, *Proc. Phys. Soc.* **50**, 345 (1938).

⁶ J. A. Becker, *Phys. Rev.* **34**, 1323 (1929).

⁷ H. Gaertner, *Phil. Mag.* **19**, 82 (1935).

⁸ H. Huber and S. Wagener, *Zeits. f. tech. Physik* **23**, 1 (1942).

⁹ J. A. Darbyshire, *Proc. Phys. Soc.* **50**, 635 (1938).

¹⁰ J. A. Darbyshire, *Proc. Phys. Soc.* **50**, 964 (1938).

* Airco "Dry Nitrogen," 99.6 percent pure, passed through a H₂SO₄ bubbler and two CaSO₄ drying towers.

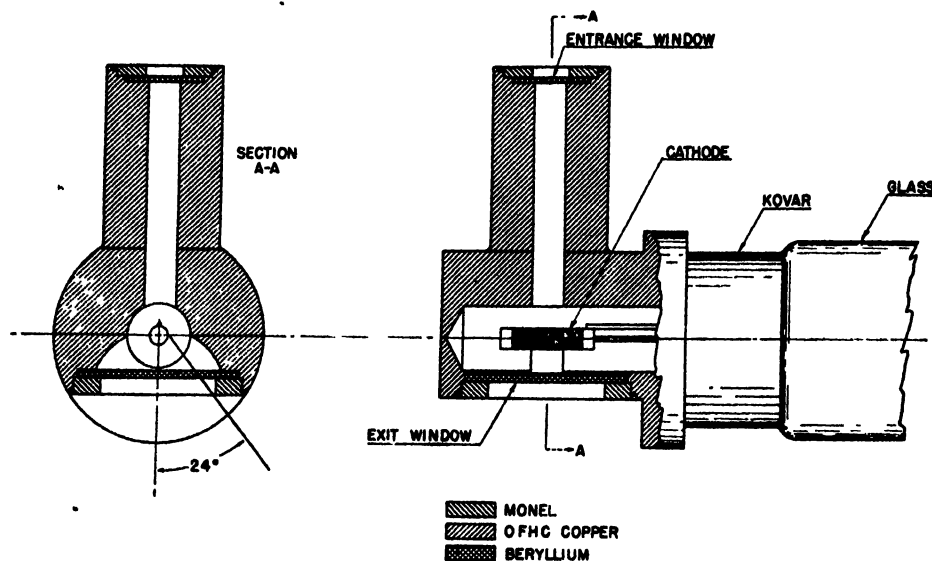


FIG. 1. Anode block for x-ray diode.

box by the wrist bands, permit this operation. After cracking the tube, the cathode is dipped in a protective coating, several of which were investigated. Polystyrene in a CCl_4 solvent gave adequate protection during periods of low humidity and was superior to nitrocellulose preparations. A special protective wax** coating was found to preserve the oxides without detectable chemical change for several days and in some instances for as long as three to four weeks. The use of the latter was generally adopted. Two weak diffraction lines are obtained from the wax corresponding to spacings of about 4.1 Å and 3.7 Å. Finally, the coated cathode was dipped in a suspension of tungsten powder in a nitrocellulose binder to provide a standard crystal size for calibration purposes.

Figure 1 shows the anode block of the special diode designed to study line changes occurring in a single cathode. The large beryllium*** exit window provides for the examination of diffracted rays up to 24° on either side of the incident beam. Primary radiation scattered from the entrance window is effectively checked by the built-in collimator tube, while scattering from the exit window is stopped by a strip of lead placed just outside the window. The windows, collimator, and Kovar ring were soldered to the block with a

Ag-Cu eutectic alloy, without flux, in a H_2 atmosphere. Two diode types were used: in the first (see Figs. 1 and 2) the cathode is arranged to slide from its position between the windows to the upper part of the block where it is centered in a cylindrical anode section for operation and testing. In the second diode type, the cathode is mounted rigidly on the glass press which is separated from the anode block by a ground glass connection. This type is pumped continuously; it is not designed for emission studies but is used only to determine the effect of temperature and time on coating materials.

Since most of the cathodes which were subjected to diffraction analysis had an outside diameter of 3 mm, a 3-mm $\times$ 0.5-mm collimator was used to provide a beam of nearly parallel x-rays incident over the semi-periphery of the cathode, Fig. 3A. Cameras of radius 4.70 cm were used in this study. As the sample was opaque to the characteristic radiation (filtered $\text{CuK}\alpha$ x-rays),* the diffraction pattern arose from conditions not generally utilized in x-ray analyses. Figure 3B shows a cathode of radius R , bathed in a parallel beam of radiation, scattering diffracted rays at an angle 2θ with the incident direction. Because of the very high absorption of the coating and base metal, only those portions of the periphery contained in S and S' can contribute

* A low softening temperature wax of the cerasin type.

*** Beryllium windows obtained from the Machlett Laboratories, Inc. Effective apertures and thicknesses, $\frac{1}{4}'' \times .020''$ and $\frac{1}{4}'' \times .040''$.

* Machlett type A-2 diffraction tube operated at 35 kv peak.

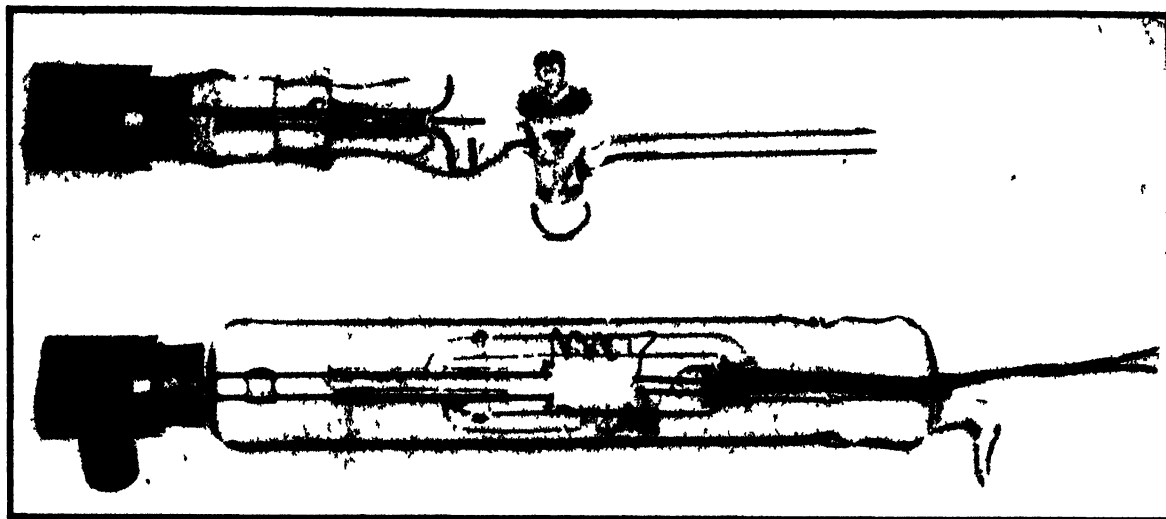


FIG. 2. Two types of x-ray diodes showing glass construction.

to the scattering AB and $A'B'$. For 2θ less than 90° , the diffracted rays are produced in two discrete areas so that inhomogeneities on the two sides of the irradiated cathode surface may be recognized by differences in the diffraction pattern on the two sides of the film.

Several disadvantages are encountered in the use of large diameter cathodes for x-ray studies which were avoided by previous workers³ who used thin coated wires. The wide collimator necessary to completely bathe the sample may permit the use of a large target focal spot area and produce blurred diffraction lines. However, by setting the collimator slit perpendicular to the target face and adjusting for the radiation at a small angle, sharp lines are obtained. The large

sample diameter introduces an appreciable error in the diffracted line position as recorded on the photographic film. This is a non-linear correction which depends critically upon the centering of the cathode within the camera as well as on the shape of the sample so that the usual correction¹¹ is not strictly applicable. From known diffraction lines recorded on each film from the protective wax, tungsten powder, and in some cases, the nickel base metal, a correction curve is obtained and applied to each unknown line position of the pattern.

III. CONVERSION STUDIES

That step in the processing of a vacuum tube in which the carbonate cathode coating is con-

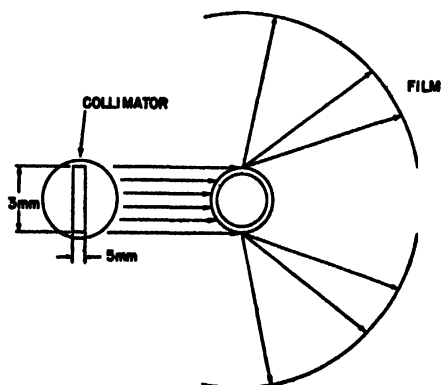


FIG. 3. (A). Slit collimator for bathing 3-mm cathode with incident beam.

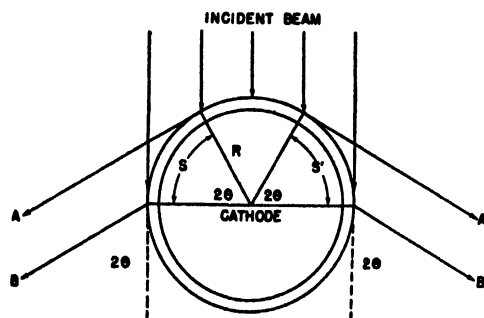


FIG. 3. (B). Showing effective scattering segments S and S' for opaque sample.

¹¹ R. Glocker, *Materialprüfung mit Röntgenstrahlen* (Verlagsbuchhandlung, Julius Springer, Berlin, 1936), p. 160.

verted to the oxide is generally considered critical and difficult to control. It is, indeed, this operation and its relationship to previous and subsequent treatment of the cathode materials that gives rise to the so-called "art of cathode making," particularly, when optimum thermionic emission is desired.

Benjamin, *et al.*,⁵ reported the dependence of conversion rate and d.c. emission on the method of carbonate preparation which presumably allows variation in only the particle size. Thus, higher emission values were associated with small particle size carbonates which they believed convert to oxide particles of the same size. Although conversion and activation studies of mixed carbonates, $\text{BaCO}_3 + \text{SrCO}_3$, are described^{1,4} in some detail, a similar investigation of the conversion of the solid solution, $(\text{BaSr})\text{CO}_3$, is not reported. Such a study should be of practical interest since solid solution carbonates are now used extensively in commercial vacuum

activation of the oxide as well as decomposition of the carbonate.

The conversion of an equal molar barium-strontium solid solution carbonate,* $(\text{BaSr})\text{CO}_3$, was studied by means of a series of x-ray diffraction patterns. Each of a series of identical cathodes** was partially converted and stopped at a different point in a standard conversion schedule. Each cathode was mounted in a glass bulb which was evacuated by means of a three-stage oil diffusion pump. Following a 60-minute bake-out at 450°C , cathodes were heated to about 700°C and held at that temperature until the carbon residue from the nitrocellulose coating binder was no longer visible. The heater current was then increased in regular steps, so chosen that complete decomposition would take place in about 15 minutes. Finally, each cathode was prepared for x-ray examination using the wax technique previously described.

Table I shows temperature and pressure values recorded at the corresponding times the conversion schedule was interrupted and represents, therefore, a complete cathode conversion. Seven cathodes were given additional heat treatments indicated in the lower section of Table I. Each time and temperature listed represents the total treatment following the above decomposition schedule. Temperatures listed are based on uncorrected pyrometer readings made at the center of the cathode from which point the x-ray diffraction patterns were also taken. Values below 750°C were obtained from an extrapolated heater wattage *vs.* temperature plot. An ionization gauge and d.c. amplifier with a maximum amplification of 1000 was used to measure the pressure in the vacuum system.

Seven of the x-ray diffraction patterns listed in Table I are reproduced in Fig. 4A, B, C, E, F, G, I, and J. Pattern A of the coating, following a 30-minute bake-out at 450°C , is identical with the original carbonate pattern except that scattering from the nitrocellulose binder is now missing. The small crystal size of the carbonate coating is reflected in the diffuse nature of the diffraction "lines." All strong "lines" with the

TABLE I. Conversion schedule and heat treatment of $(\text{BaSr})\text{CO}_3$ cathodes.

Time from previous step	Maximum temperature	Pressure	Pattern	Results
60 min.	450°C	10^{-7} mm	A	No physical or chemical change in carbonate coating
1	600	10^{-6}		No change
3	700	5×10^{-4}	B	Growth in carbonate crystal size
2	740	$> 10^{-3}$	C	Crystal growth, SrO lines present
2	795	$> 10^{-3}$	E-F	BaCO_3 and SrO on one side, BaO and SrO on opposite side
4	810	$< 10^{-6}$	G	BaO and SrO
30	Subsequent heating:			decomposition
120	800	10^{-7}	I	$(\text{BaSr})\text{O}$
5	800	10^{-7}		$(\text{BaSr})\text{O}$
5	850	10^{-7}		$(\text{BaSr})\text{O}$
5	890	10^{-7}		$(\text{BaSr})\text{O}$
5	940	10^{-7}		$(\text{BaSr})\text{O}$
5	1000	10^{-7}		$(\text{BaSr})\text{O}$, large crystals
5	1050	10^{-7}	J	$(\text{BaSr})\text{O}$, very large crystals

tubes. In this report the conversion process shall refer to those steps in the processing of an oxide cathode which occur prior to the actual drawing of emission current and shall include, therefore, what has been referred to in the past as thermal

* Coating Type C51-2, obtained from the Raytheon Manufacturing Co., Newton, Massachusetts.

** The cathodes were coated approximately 10 mg/cm^2 , 8-mm coated length, on $3 \times 14 \text{ mm}$ indirectly heated, Grade "A" nickel sleeves.

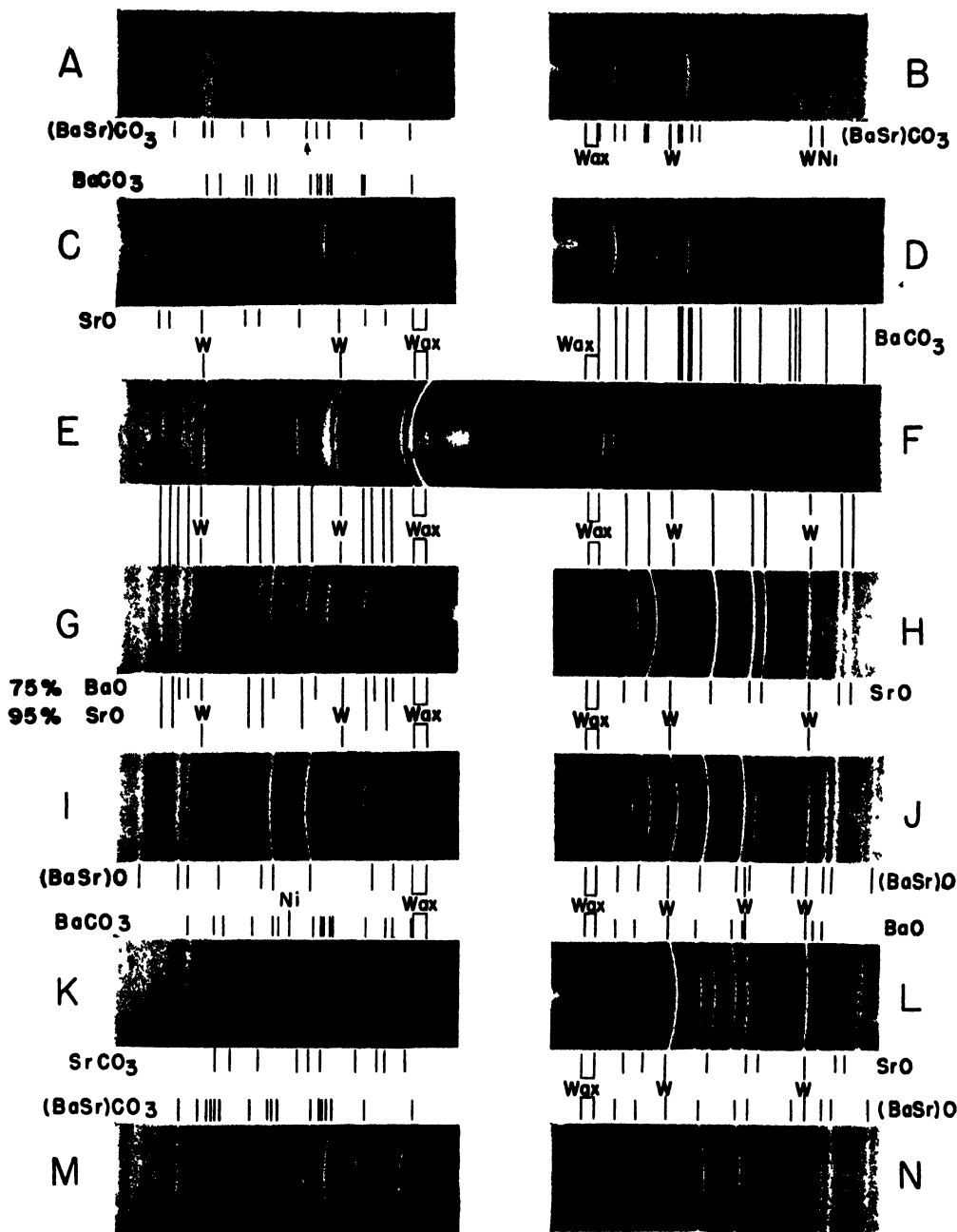


FIG. 4. X-ray diffraction patterns of cathode samples. See text.

exception of the one marked by an arrow are multiple diffraction lines. This may be seen by close inspection of *B* or *M* which are patterns of larger crystals of the same material. The origin of this multiple structure is not the presence of two or more solid solution phases since pure BaCO_3 in *D* shows a similar multiple line pattern. Attention is called to this point, for although the

multiple nature of some of these lines is recognized,¹² they are frequently¹³ treated as single reflections. Line breadth measurements for the determination of crystal size may thus be correctly applied only to the single line indicated. It is evident from the somewhat sharper lines in

¹² J. M. Cork and S. L. Gerhard, *Am. Min.* 16, 71 (1931).

¹³ A.S.T.M. x-ray analysis card file.

pattern *B* of the cathode after 3 minutes at 700°C that crystal growth has begun although no chemical transformation has occurred. Pattern *C*, taken after an additional 2 minutes at 740°C, shows the beginning of the decomposition process. Lines of both BaCO₃ and SrO are present indicating the decomposition of SrCO₃ at this stage. Figure 4E and F shows an interesting diffraction pattern of a cathode at the 795°C stage. One side of the cathode gave lines only of BaCO₃ and SrO while the opposite side, which was probably at a slightly higher temperature, had completely converted to BaO and SrO. Solid solution formation had begun since two phases are present, 95 percent SrO - 5 percent BaO and 75 percent BaO - 25 percent SrO. (To aid in the identification of these compounds, compare with *D* of pure BaCO₃, *H* of pure SrO, and *L* of mixed crystals of BaO and SrO.) Inhomogeneities over the cathode surface frequently appear during this stage. Because of the extreme temperature sensitivity of these processes, composition variations are often encountered along the length of the coating. Pattern *G*, taken at the completion of the decomposition process at 810°C, also shows lines of the phases 95 percent SrO and 25 percent SrO. Subsequent heat treatment (Table I) shows that a homogeneous solid solution is formed in 30 minutes at 800°C, Pattern *I*, or in 5 minutes at 850°C.

For convenience in the further discussion of the conversion process a division into three separate stages will be made: (1) crystal growth in the carbonate prior to decomposition, (2) decomposition and solid solution formation, and, (3) crystal growth in the oxide following decomposition.

Decomposition and Solid Solution Formation

Although the presence of lines of both BaCO₃ and SrO in Fig. 4C and F shows the complete decomposition of the SrCO₃ component of the solid solution to SrO before the conversion of BaCO₃ reaches appreciable proportions, the exact mechanism of the process cannot be unambiguously described. Two mechanisms are suggested: (1) the dissociation of the carbonate solid solution into its separate components just prior to the conversion of SrCO₃, or (2) the conversion to an oxide solid solution and the immediate recombination of BaO in the conversion gasses.

Since the conversion products of an initial solid solution carbonate were mixed crystals of BaO and SrO, the conversion study was extended to mixed carbonates which were reported^{1,4} to yield the same conversion products. Pattern 4-K is of a commercial mixed carbonate coating, RCA 33C-133. This coating, when decomposed at 800°C, using a conversion schedule similar to that for the solid solution carbonate, Table I, gave pattern 4-L. Lines of BaO and SrO mixed crystals are present. The rate of solid solution formation was investigated in a series of cathodes flashed, after decomposition, for 5 minutes at temperatures of 860°C, 900°C, 1000°C, 1050°C, and 1100°C. Only those patterns of cathodes flashed at 1000°C and above showed single lines characteristic of a solid solution. This relatively slow rate of solid solution formation is in agreement with the findings of others who report^{1,4} that formation from mixed carbonates requires up to an hour at 925°C. Broad lines were found in these patterns which were interpreted as representing a range of solid solutions of near equal molar composition rather than as an indication of small crystal size.

It appears then that in the conversion of mixed (BaCO₃+SrCO₃) and solid solution (BaSr)CO₃, there is an increased rate of solid solution formation and a greater homogeneity in the final oxide solid solution in the case of (BaSr)CO₃. A comparison of *G* and *L*, both mixed oxides which were decomposition products of (BaSr)CO₃ and (BaCO₃+SrCO₃), respectively, showed no difference in crystal structure other than a possible variation in crystal size.

Oxide Crystal Size

Line breadth measurements made from microphotometer tracings were used in the conventional manner¹⁴ to determine the average crystal size of carbonate and oxide samples. Corrections¹⁵ for instrumental broadening effects were made from nearby large crystal tungsten lines. By means of this method the crystal size of both phase components of pattern *G* were found to lie between 135 and 180A, whereas, those of pattern *L* are in the range 240 to 325A. Further heating at

¹⁴ M. v. Laue, *Zeits. f. Krist.* **64**, 115 (1926).

¹⁵ B. E. Warren, *J. Am. Ceram. Soc.* **21**, 49 (1938).

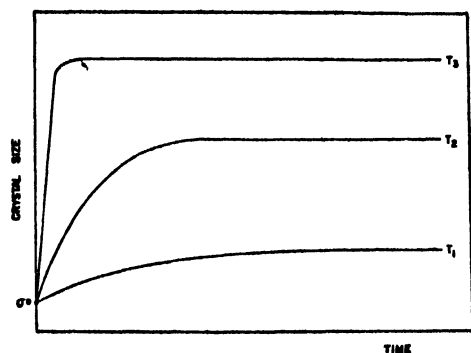


FIG. 5. Idealized crystal size growth curves of carbonates and oxides of initial size σ_0 growing at temperature $T_1 < T_2 < T_3$.

temperatures below 940°C did not result in appreciable crystal growth although solid solution formation did occur. Pattern *I* of the single oxide phase showed an average crystal size of 180A and after 2 hours at 800°C the size was 190A. Five minutes at 940°C gave a crystal size of 220A. Rapid growth occurs at 1050°C since the lines of pattern *J*, a cathode flashed for 5 minutes at this temperature, showed a decided "graininess" which is indicative of crystal sizes in excess of 10^{-3} cm. This suggests that the difference in crystal size of the mixed oxides, Fig. 4G and L, might be responsible for the observed differences in rate of formation and initial homogeneity of the final oxide solid solution. This seems to be a reasonable explanation since the formation of a homogeneous solid solution without crystal growth requires the interchange of at least half the molecules of any one crystal with adjacent crystals, a diffusion process which is expected to be strongly dependent on crystal size and on temperature in the manner indicated.

Carbonate Crystal Size

Since it is known that growth occurs in many crystals at temperatures above 400°C, and a correlation between emission and carbonate particle size is reported,⁴ it was suspected that certain cathode conversion techniques might be associated with just such a particle growth to an optimum size. Three commercially prepared cathode coatings, (1) an equal molar (BaSr)CO₃,* (2) BaCO₃,** and (3) SrCO₃** were heat treated

* Baker No. 3 Radio Mixture.

** Obtained from Raytheon Manufacturing Company.

in a muffle furnace at temperatures of 450°C to 1050°C for times of from 5 minutes to 7 hours. These dry coatings were held in an enclosed nickel boat during treatment. X-ray diffraction patterns and photomicrographs were taken to measure both crystal and particle size of these samples. It should be noted that, although in some cases the crystal and particle sizes of a sample are identical, particles usually consist of a conglomerate of crystals, the size of which may be smaller by several orders of magnitude. Many of the samples heat treated above 600°C had crystal sizes in excess of 10^{-3} cm; hence, only semi-qualitative observations could be made based on the spot size¹⁶ or on the degree of "graininess" of the diffraction lines. With heat treatment above 500°C, the crystal sizes of all three coatings were found to grow and to reach an equilibrium size which depended primarily on the temperature. This is a behavior which was observed in the growth of magnesium oxide¹⁷ crystals as well as carbon black particles.¹⁸ Idealized growth curves are shown in Fig. 5. Both the equilibrium size and growth rate are seen to increase with temperature. At 500°C all three coatings were found to increase from an initial crystal size of about 250A to greater than 10^{-3} cm in several hours, whereas at 600°C, a similar growth occurs in about 15 minutes. Still larger crystals were grown at higher temperatures by heating in an atmosphere of CO₂ to prevent decomposition.

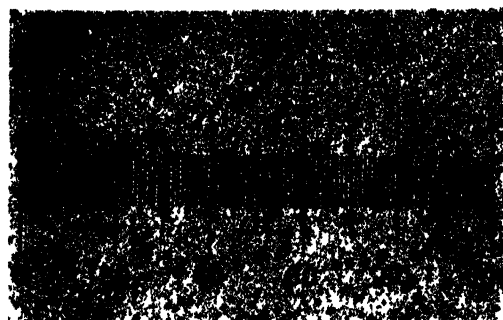
The simultaneous growth of particle size may be seen in Fig. 6 which shows photomicrographs of BaCO₃ initially and after heating at 500°C for 15 minutes and 4 hours. A microscope magnification of 440X gives an over-all enlargement of 1.5 microns per division. Table II lists the minimum, average, and maximum size of the particles before and after treatment at 500°C and 600°C.

Thus, it seems that the average crystal size of untreated carbonates is in some cases less than 1/100 of the measured particle size; however, after growth (see Fig. 6), the particles appear to be single crystals. Particle growth at 600°C takes place about three times as rapidly as is observed

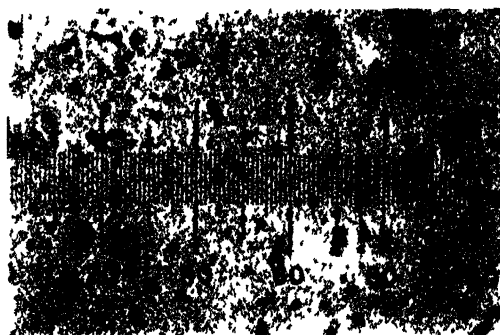
¹⁶ G. L. Clark, *Applied X-Rays* (McGraw-Hill Book Company, Inc., New York, 1940), p. 506.

¹⁷ J. A. Neuendorffer, M.I.T. Thesis (1941).

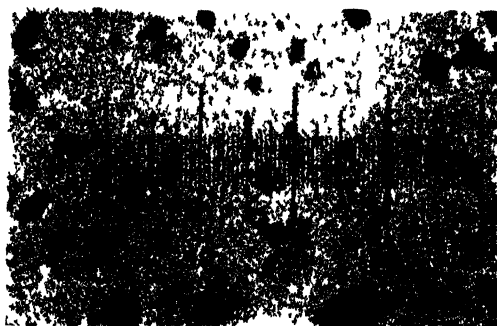
¹⁸ J. Biscoe and B. E. Warren, *J. App. Phys.* 13, 364 (1942).



(A)



(B)



(C)

FIG. 3. Growth in particle size of BaCO_3 induced by heat treatment. (A) Initial carbonate, (B) 15 min. at 500°C , (C) 4 hr. at 500°C . Magnification 1.5 microns per division.

at 500°C . Samples heated at 450°C showed only a slight increase in size.

Attempts made to influence the oxide crystal size by pretreatment of the carbonate or by variations in the conversion schedule were not successful. As with the carbonate, the crystal size of the oxide approached an equilibrium value which depended primarily on the temperature. In the temperature range of crystal growth this

equilibrium size is reached in a matter of minutes so that the oxide crystal size can be expressed solely as a function of the maximum cathode temperature reached at any time during life. Cathodes converted at 800°C and flashed momentarily at 1000°C have crystals of a size characteristic of those converted at 1000°C , Fig. 4J.

A further indication of the unrelated nature of the carbonate and oxide crystal size is seen by comparing $(\text{BaSr})\text{CO}_3$ heated to 800°C in CO_2 to prevent conversion and $(\text{BaSr})\text{O}$ converted from this large crystal carbonate at the same temperature in vacuum. The carbonate equilibrium size appears to be at least an order of magnitude larger than the equilibrium size of the oxide at the same temperature, Fig. 4M and N.

No evidence was found to indicate a change in crystal size during the normal life of an oxide cathode provided that the initial maximum cathode temperature is not exceeded.

It has been shown that (1) both the carbonate crystal size and particle size begin to increase at about 500°C and approach equilibrium values which depend on the time and temperature of treatment, (2) a similar crystal growth in the oxide solid solution begins at about 950°C , and (3) the oxide crystal size depends primarily on the maximum temperature of treatment above 950°C and bears no relationship to the crystal or particle size of the carbonate from which it was formed. It is probable, therefore, that no correlation exists between carbonate particle size and oxide particle size. Thus, it is difficult to see how the emission can depend upon the carbonate particle size in a fundamental manner.

IV. THERMAL EXPANSION COEFFICIENTS

The adherence of coating to base metal is an important parameter in the study of the thermionic emission of a variety of oxide coating, base metal combinations. For example, BaCO_3 on a chrome-plated nickel sleeve converted to BaO in vacuum at a relatively low temperature was found to fall from the sleeve on cooling, suggesting a possible difference in the coefficients of expansion of BaO and Cr . It was noted, however, that if this cathode was heated to a high temperature, the adherence was improved. A similarly poor bonding between base metal and coating has been observed in the case of ThO_2 on W .

Values of the thermal coefficients of expansion of most of the alkaline earth oxides not reported were determined in the following manner. The oxide or carbonate to be tested was sprayed on a cathode sleeve and mounted in a continuously pumped, demountable x-ray diode (Fig. 2). The diode was rigidly mounted with respect to the x-ray tube, collimator, and camera and sealed to a high vacuum system. With this arrangement one x-ray diffraction pattern was taken of the oxide at room temperature and a second at some elevated temperature, usually 875°C. The fractional change of lattice constant $\Delta d/d$ with temperature ΔT is simply¹⁹ the coefficient of

TABLE II. Particle growth in carbonate at 500°C and 600°C.

Time	BaCO ₃			SrCO ₃			(BaSr)CO ₃		
	Min	Av	Max	Min.	Av.	Max	Min	Av	Max
0 min.	0.5 μ	2 μ	4 μ	0.5 μ	2 μ	4 μ	1 μ	3 μ	9 μ
500°C									
15 min.	1	3	6	1	3	5	1.5	3	6
4 hr.	1.5	4	8	2	5	11	2	5	10
600°C									
5 min.	1.5	2.5	7	1	3	5	1.5	5	12
15 min.	1.8	3	7	1.5	4.5	12	2	9	20

expansion for crystals having a cubic structure. With this method the lattice constant can be measured to an accuracy of 0.2 percent and since the change in this parameter is about 1.5 percent at 875°C for BaO, an over-all accuracy of about 15 percent is indicated.

Values of the linear coefficient of thermal expansion thus obtained for the various coatings

tested are shown in Table III. Reported values for metals which are used for base metals are also shown in this table. Two factors should be considered in attempting to correlate expansion coefficients and bonding properties: (1) the values given in Table III described the true crystal expansion and represent, therefore, the maximum

TABLE III. Linear expansion coefficients of oxide coatings and base metals.

Material	$\Delta d/d\Delta T \times 10^6$
SrO	32.5 $\pm$ 2.6
(BaSr)O	26.6 $\pm$ 3.0
BaO	17.8 $\pm$ 2.1
ThO ₂	11.3 $\pm$ 0.4
Nickel	13.5
Platinum	9.0
Copper	16.1
Tungsten	4.6

value that would be encountered in a low density sprayed coating, and (2) if an interface layer of different composition²⁰ is present between the coating and base metal, the adherence will depend on the interface coefficient as well. It is interesting to note, however, that the oxides have larger coefficients than their respective base metals so there is a gripping action of the coating on cooling.

No structural change other than that caused by the lattice expansion was observed in the patterns of the coatings at 875°C.

The author is indebted to Dr. E. A. Coomes and members of the Cathode Research Group for valuable suggestions and assistance during the course of this investigation. The photomicrographs in Fig. 6 were kindly prepared by Mr. C. D. Prater of the Bartol Research Foundation.

¹⁹ A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Co., Inc., New York, 1935), p. 682.

²⁰ H. D. Arnold, *Phys. Rev.* 16, 70 (1920); A. Gehrts, *Zeits. f. tech. Physik* 11, 246 (1930).

Methods for Betatron or Synchrotron Beam Removal

EUGENE C. CRITTENDEN, JR.* AND WILLIAM E. PARKINS
Radiation Laboratory, University of California, Berkeley, California

(Received January 9, 1946)

Two methods for bringing out a partially collimated electron beam from a betatron or synchrotron are discussed. Both of the methods become easier as the size of the accelerator is increased. One method makes use of a perturbing magnetic field to focus the electrons as they are made to leave the field of the accelerator by means of orbit expansion. The other method makes use of a pulsed deflecting system where the deflecting field is applied during a time short compared to the period of revolution of the electrons.

TO date devices of the betatron and synchrotron^{1,2} type for accelerating electrons to high energies have not been adapted to bringing the electrons out in a beam for study, but rather have made use of the x-rays produced by the electrons striking some internal target. In the following, two methods are described which may be used to eject a partially collimated electron beam from such devices. These methods are made practical by the increased size of the present largest instruments in use or under construction. No experimental test has been made of the proposals, but the current importance of the beam

removal problem has prompted submitting them without awaiting construction of a suitable apparatus at the Radiation Laboratory. The methods are suggested in connection with the removal of an electron beam although they may be equally useful when applied to ion beams, as in a frequency modulated cyclotron. They may be employed to advantage in any accelerating device using a magnetic field to restrain charged particles to move in cyclic paths such that the radial displacement from one turn to the next is normally small.

The magnetic field at any instant of time is usually expressed as:

$$H = H_0(r_0/r)^n,$$

where r is the distance from the center of the system, H_0 is the magnetic field at some fixed distance r_0 from the center of the system, and n is a dimensionless constant. At the edges of the pole face the field varies in a manner to which this form is not well suited since H_0 , r_0 , and n must all be considered variable. However n , represented as the negative of the slope of a $\log H$ versus $\log r$ plot, remains a useful quantity.

As the electron orbits are slowly expanded, as in the betatron by means of expanding coils, or in the synchrotron by means of turning off the oscillator at a time after the occurrence of peak magnetic field, the electrons will ultimately reach a value of r at which their orbits become unstable. The solid curve in Fig. 1 is typical of the field near the pole face edge in the median plane of the gap. Here instability occurs at point A, where $n=1$, if the electron has zero radial oscillation amplitude about its instantaneous orbit. If the electron has an appreciable radial oscillation

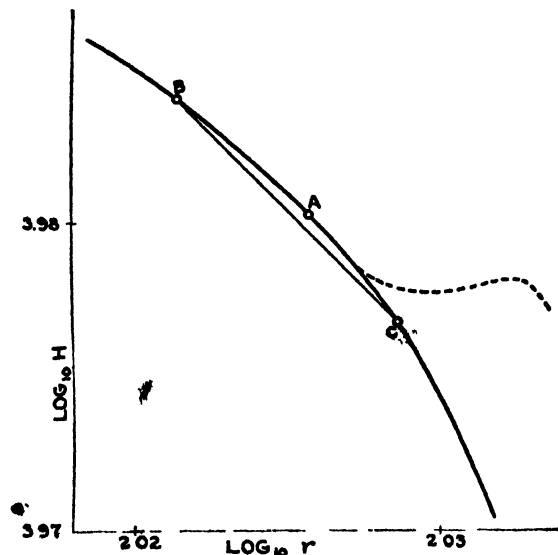


FIG. 1. Magnetic field to be expected at the edge of a synchrotron magnet of 1-meter orbit radius, based on data from a $\frac{1}{2}$ scale direct current model magnet. The dotted line represents the field as perturbed by the coil of Fig. 3a with a current of 510 amperes.

* On leave from Case School of Applied Science, Cleveland, Ohio.

¹ E. M. McMillan, Phys. Rev. 68, 143 (1945).

² V. Veksler, J. Phys. U.S.S.R. 9, 153 (1945).

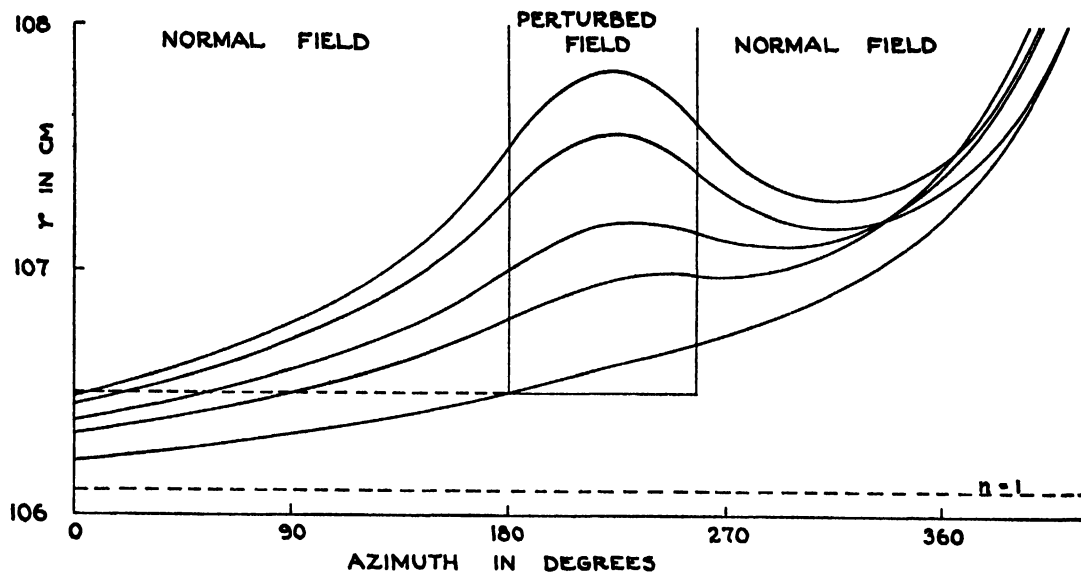


FIG. 2. Predicted electron trajectories for the magnetic field of Fig. 1 with a perturbed field given by the dotted line of Fig. 1 over 78° of azimuth.

amplitude, it will be released somewhat earlier in time. This condition is illustrated in Fig. 1 by an electron which has its instantaneous orbit at *B*, but by virtue of radial oscillation reaches a point of instability, *C*, determined by the intersection of the magnetic field curve and a straight line of slope -1 passing through *B*. After release the electrons move outward with increasing radial velocity. If allowed to continue, they leave the magnetic field of the accelerator producing a sheet of energetic electrons spread uniformly through 360° of azimuth.

The shape of magnetic field ordinarily encountered in a betatron or synchrotron is such that after an electron has passed a value of *r* about $\frac{1}{2}$ percent greater than *r* at the edge of instability, its value of *r* increases another one percent in traveling about 180° of azimuth. In a large accelerator this radial velocity is a convenient one for handling the released electrons in a focusing field. If the magnetic field is altered over a small range of azimuth in a manner similar to that represented by the dotted line in Fig. 1, the electrons can be made to emerge as a partially collimated beam. The electrons could be similarly handled in an electric field but proper shaping of an electric field seems to have more practical difficulties.

To illustrate a focusing scheme, results of calculations on one arrangement planned for a

particular application will be described here. A perturbed field corresponding to the dotted line of Fig. 1 was employed over a range of 78° in azimuth. The electron trajectories were determined by the graphical method described in the paper immediately following. Figure 2 shows the trajectories for electrons of zero radial oscillation amplitude. The unperturbed magnetic field employed was that observed in a $\frac{1}{8}$ -scale model of the magnet for a synchrotron of 1-meter radius. The perturbing field producing the focusing effects was the calculated field produced by the array of conductors shown in Fig. 3a where each conductor carries 510 amperes. The wires were arranged in a plane to permit construction by firing silver conductors into grooves in flat ceramic plates, thus assuring good mechanical support. The particular array was chosen to give negligible perturbing field inside the region of stability and to have a field profile giving small aberration, i.e., focusing effects too strong or too weak for electrons entering the perturbed region at the average radial displacement. This array of conductors gives a usable field depth of 1.0 cm in the axial direction. To avoid excessive values of the average power dissipation the coil need only be pulsed for a few microseconds when the electrons are expanding over the edge of instability. The use of local iron to produce the focusing field would avoid the complication of an

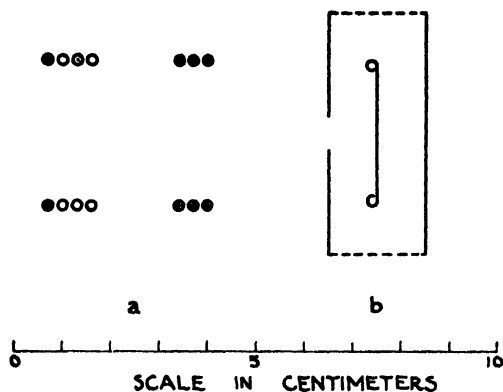


FIG. 3. (a) Cross section of the conductor array used to produce the perturbing field. Numbering from left to right, wires 1, 5, 6, and 7 carry current into the page, wires 2, 3, and 4 carry current out of the page. (b) Cross-section of a deflecting electrode. The center of the magnet lies to the left for both diagrams.

electric pulsing circuit but has difficulties with regard to mechanical support, production of a correctly shaped field, and flexibility of adjustment. The axial defocusing inherent in the region of negative n occurring over a part of the dotted line of Fig. 1 is not serious because of the small path length of an electron in this field.

Figure 2 indicates that all trajectories originating in a range of azimuth of 174° would be brought into a beam crossing $r=108.0$ cm within a range of azimuth of 16° . This beam would consequently contain 48 percent of the accelerated electrons. Tracing these rays out to where the field is zero predicts a beam with an angular divergence of 17° . In this example the inner edge of the region of perturbed field was placed at $r=106.5$ cm whereas $n=1$ at $r=106.1$ cm. This was done to avoid introducing a perturbation in the orbits of radially oscillating electrons before they reached their edge of instability. Electrons which have their edge of instability at $r=106.5$ cm would have a radial oscillation amplitude corresponding to a beam 1.5 cm wide in a region where $n=\frac{3}{4}$. These electrons are properly focused by the same coil but a current of 440 amperes is required in place of the 510 amperes for non-oscillating electrons. Since radially oscillating electrons are released early and at a time dependent on their amplitude, a properly varying current in the focusing coil could serve for all electrons.

If the maximum amplitude of radial oscillation proves to be less than the value assumed, which

seems likely, the edge of the perturbing field could be moved closer to the circle at which $n=1$. This would increase the fraction of the accelerated electrons appearing in the beam. An upper limit to the improvement attainable would be reached when the electrons began to pass through the outer edge of the perturbing field on the second revolution.

A modification which has some possibility for increasing the fraction of the electrons appearing in the beam, but that is difficult to evaluate without actual trial, is the application of a field variation in the region of the expanding orbits, over a small azimuthal range. The disturbance produced in the orbital motion during expansion might be such as to cause the electrons to exhibit a preferential range of azimuth in which they would cross their edge of instability. If this range were so located as to lie within the range from which electrons are successfully focused, an increase in the fraction of the electrons in the beam would result.

The second method to be discussed for producing a collimated beam of electrons is in effect an extreme case of concentrating the range of azimuth in which electrons pass into the region of instability. For large accelerators it seems probable this method could produce a fairly well collimated beam without the help of an auxiliary focusing field. By allowing the orbits to expand almost to their edge of instability, deflection of the entire beam could be accomplished with a reasonable electric or magnetic field, provided this field could be applied in the form of a flat topped pulse, rising in a time short compared to the period of the electrons. The fraction of the total number of electrons brought into a collimated beam would depend on the degree to which the electrical problems could be solved.

As an example, one possible solution would be the use of a charged transmission line of electrical length at least half the circumference of the accelerator, connected by means of spark discharge to an electric deflecting plate. The deflecting plate should resemble as nearly as possible a short continuation of the transmission line and be connected at its extreme end to a terminating resistance equal to the characteristic impedance of the transmission line. The time of firing of the spark is not critical and breakdown times of the

order of 10^{-9} second are known to be possible. If employed on an accelerator having a 1-meter orbit radius, where the period of the electrons is 2×10^{-8} second, this method should yield a reasonable fraction of the electrons in a collimated beam.

To illustrate the degree of collimation produced by such a deflector, consider the electrode structure shown in Fig. 3b. Such an arrangement should be constructed to spiral outward at the rate required to keep the deflected beam nearly grazing the inner electrode. With +80 kv applied to the transmission line, approximately +40 kv would appear on the inner electrode, and would give a fairly uniform electric field of 37,000 volts per cm in the region just to the left of the inner electrode. If the electrode is so connected that the conventional current flows in the direction opposite to the direction of motion of the electrons to be deflected, the magnetic field due to the current assists the deflection and is quite uniform in the region occupied by the beam. In the geometry considered here the force on a beam electron due to the magnetic field produced by current in the deflector will be almost 40 percent of the total deflecting force. For the deflecting electrode occupying 40° of azimuth and in the unperturbed magnetic field given by the solid line of Fig. 1, suppose the deflecting potential is applied at a

time when the electron beam has expanded until its center lies 0.6 cm inside the circle at which $n=1$. In this case, considering both electric and magnetic fields produced by the deflector, graphical ray tracing predicts an angular divergence of 18° in the emergent beam for an internal beam width of 0.5 cm at the deflector. This corresponds to a beam 0.4 cm wide in an $n=\frac{3}{4}$ region. The angular divergence decreases approximately linearly with internal beam width and decreases with increasing applied voltage.

The foregoing has pointed out how electron beams might be withdrawn from large accelerators of the betatron or synchrotron type. One scheme uses a magnetic lens to concentrate electrons leaving with 360° spread in azimuth, the other employs a deflector which starts the outward motion in a small range of azimuth. At present the lens appears to be a slightly more effective and practical solution. If beam intensity and collimation were prime considerations, it is conceivable a combination of deflector and lens would comprise the best arrangement. These methods are not restricted to use with an electron beam but may be valuable when employed with an ion beam, such as is produced in a cyclotron.

The authors wish to acknowledge the information and advice given in connection with this problem by Professor E. M. McMillan.

A Graphical Method for Determining Particle Trajectories

WILLIAM E. PARKINS AND EUGENE C. CRITTENDEN, JR.
Radiation Laboratory, University of California, Berkeley, California
(Received January 9, 1946)

A graphical ray tracing scheme is described which gives improved accuracy for tracing particle orbits of the type encountered in the cyclotron, betatron, and synchrotron. The method allows a large increase in scale factor by elimination of the necessity of a compass arm. The motion of the particle is deduced from the motion of a working point, related to the instantaneous center of curvature of the particle orbit.

AT present there is an increasing number of applications for controlling the motion of charged particles by electric and magnetic fields. Except in simple cases the problem of determining trajectories is difficult. A graphical method is described here which is applicable in instances

where the motion is essentially confined to a plane and where the radius of curvature of the particle is known as a function of space coordinates. In particular the method is useful for problems of motion of charged particles in devices similar to the cyclotron and betatron.

A graphical method in common use employs step-wise construction of the trajectory. A compass is set for a radius corresponding to the initial radius of curvature of path and placed on a layout of the coordinate system so that the compass arm extends normal to the initial path direction. Then the compass is swung through some angle which would bring the particle to a coordinate position where its radius of curvature is slightly changed. Here the compass arm length is changed correspondingly by moving its center without changing the arm position. Tracing may be continued by repeating the process of drawing a small arc and adjusting to a new radius of curvature. By taking sufficiently small steps any desired accuracy may be obtained within the mechanical limits of error of the graphical technique. But even this limitation may be reduced by using larger scale drawing instruments, and the method only breaks down when the accuracy required makes it too cumbersome to be practical.

In order to eliminate the cumbersome feature, the compass arm, a new system has been devised which essentially deduces the trajectory from the motion of the center of curvature of the particle. In cases where the center of curvature and origin remain within a region of dimensions small compared to the radii of curvature, any plot of the motion of the center of curvature can be done to a much larger scale than the particle trajectory. Qualitatively the method depends on determination of the progress of the particle in a polar coordinate system from the motion of a "working" point which is related to the center of curvature and rotated about the origin.

In practice the method can be carried out as follows. A piece of tracing paper is placed over a piece of graph paper serving as a rectangular coordinate system. A pin is stuck through some point designated as origin and into a drawing board below permitting rotation of the tracing paper with respect to the graph paper about this center. A working point is plotted on the tracing paper with Cartesian coordinates, $x_1 = r_1 \sin \alpha_1$ and $y_1 = r_1 - R_1$, where r_1 , R_1 , and α_1 are known from the initial conditions of the particle and the properties of the field and represent, respectively, the polar coordinate radius of the particle, its radius of curvature, and the angle from the polar coordinate radius to the line connecting the

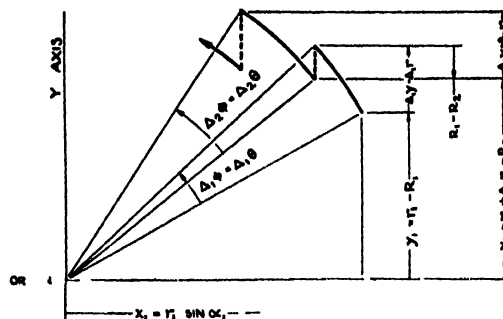


FIG. 1. Typical path of the working point shown projected through the tracing paper onto the coordinate paper.

particle and its center of curvature. A reference mark is also made to indicate the initial angular position of the tracing paper. Assuming the particle is traveling in an r, θ coordinate system, but the working point in a Cartesian system, some convenient increment of r is chosen and the tracing paper rotated until the working point moves a y distance equal to the chosen increment Δr . The proper rotation of the tracing paper is in the direction opposite to the direction of rotation of the particle. It is as though the polar system in which the particle moves rotates at the proper rate to keep the particle on the positive y axis of the Cartesian system, while one origin is common to both systems. At this new setting an angular reference mark is made to determine the change in θ . The particle can now be assumed to be at $r_1 + \Delta r$ and $\theta_1 + \Delta \theta$ where $\Delta \theta$ was the angular rotation of the tracing paper. For this new location the radius of curvature, R_2 , is computed from a knowledge of field conditions. Without moving the tracing paper a new working point is plotted with y coordinate equal to $r_2 - R_2$ where $r_2 = r_1 + \Delta r$, and x coordinate the same as the new x value possessed by the first working point. The paper is rotated again to change the y coordinate of the second working point by Δr and an angular reference mark made. It is advisable to keep the increments of r some constant and convenient value as much as possible and to number successive angular reference marks together with the working points. This process is repeated while the progress of the particle in r and θ is tabulated, to be plotted later if desired. Figure 1 shows the type of pattern which would be produced if the path of the working point were projected through the tracing paper onto the coordinate paper.

If the field is a function of θ instead of r , or of both coordinates, it may be more convenient to use fixed increments of θ . Should the field be a function of θ alone, values of R computed for each increment of θ can serve to trace a variety of particles with different initial conditions. For a field which is a function of r alone, a single table of $r-R$ values can be used for different initial conditions. The value of R computed should be for some point within the interval concerned. Half-way points are sufficiently accurate except when the particle is passing a maximum within the interval. Here interpolation is recommended.

Consideration of the errors involved leads to the following. If the working point had been located at the center of curvature, its coordinates would have been:

$$\begin{aligned}x &= R \sin \alpha, \\y &= r - R \cos \alpha.\end{aligned}$$

To introduce no errors, the change in r , θ , and α of the particle must be exactly determined by the change in x and y of the working point during rotation of the tracing paper. It can be shown this is true here for a sufficiently small rotation only if

$$\begin{aligned}d\theta &= d\phi, \\dr &= (r/R \cos \alpha) dy.\end{aligned}$$

Hence, use of the center of curvature as the working point requires that Δr be calculated from Δy by use of r , R , and α . In cases where R/r is nearly one and α nearly zero this correction is hardly worth while. A more important consideration in the choice of working point coordinates is the difficulty of calculating them for each point plotted. Since r and α remain constant and R changes each time a new working point is plotted, use of the equations for the center of curvature to locate the working point proves inconvenient.

Suppose the coordinates of the working point were

$$\begin{aligned}x &= r \sin \alpha, \\y &= r \cos \alpha - R.\end{aligned}$$

Here an exact determination of r , θ , and α can be shown to require

$$\begin{aligned}d\theta &= (R/r) \cos \alpha d\phi, \\dr &= (R/r) dy.\end{aligned}$$

This choice has the disadvantage that $\Delta\theta$ must be calculated from $\Delta\phi$ and Δr from Δy , unless R/r

is sufficiently near one and α sufficiently near zero. However a more than compensating feature lies in the fact that the x coordinate of each new working point is identical to the final x coordinate of the preceding point and the y coordinate is changed only by ΔR , the difference in R for the two intervals. If a procedure is desired which is exact regardless of the value of R/r or α , this last choice of working point coordinates is recommended. Use of sufficiently small intervals in plotting will determine a trajectory subject only to errors in graphical technique.

In the detailed description of the method the working point coordinates were given as

$$\begin{aligned}x &= r \sin \alpha, \\y &= r - R.\end{aligned}$$

The errors arising from the simplified use outlined for this choice of coordinates can be judged by comparison to the exact procedure just mentioned. The differences are $r - R$ replaces $r \cos \alpha - R$ for the y coordinate of the working point, $d\theta$ is assumed equal to $d\phi$ thus neglecting the $(R/r) \cos \alpha$ factor, and dr is assumed equal to dy neglecting the R/r factor. These differences all become negligible as R/r approaches one and α approaches zero. Errors of course are cumulative but the net error is usually reduced by R/r assuming values both greater than and less than one.

The advantage of the method discussed over the usual graphical method lies in the large increase in scale factor which can be employed in many instances. Such increased scale factors are possible when the center of curvature for the trajectory of the particle remains within a region of dimensions small compared to the radii of curvature. It is also true that under these conditions α will be small and R/r near unity contributing to the accuracy of the more rapid approximate scheme for locating and using the working points. This type of trajectory is characteristic of particle motion in the cyclotron, betatron, and synchrotron. Cases are not uncommon in which the motion permits a scale factor corresponding to an effective compass arm length of the order of 100 meters, and yet the tracing diagram is made on a sheet of notebook size paper.

The authors wish to acknowledge the assistance of Mr. Leslie L. Foldy in checking the expressions presented.

Analyses of Mixtures of Light Gases by Infra-Red Absorption

NORMAN D. COGGESHALL AND ELFANOR L. SAIER

Gulf Research and Development Company, Pittsburgh, Pennsylvania

(Received January 17, 1946)

A discussion is given of the application of infra-red absorption methods of analysis for light gases which do not obey Beer's law of absorption due to pressure broadening. Some of these gases do not obey Beer's law in the pure state or when contaminated by foreign gases. The method of analysis depends upon the nature and intensity of the pressure broadening effect of the different components in the sample upon each other. Data are presented showing the nature of some of these effects and illustrating the accuracy obtainable for certain types of analyses. The instrumentation used in routine gas analyses by infra-red in this laboratory is described.

IN the last few years there has been an intensified application of infra-red absorption spectrometry to problems of chemical analysis.¹⁻⁴ The technique is now being successfully applied to the analysis of mixtures in the gas and liquid phase and to identification and studies of molecular structure for materials in all phases. One application that has been of particular economic importance recently is gas analysis. Brattain, Rasmussen, and Cravath² published what was perhaps the first detailed procedure for multicomponent gas mixtures wherein they were interested in the four-carbon hydrocarbons. The same technique, which is based on the principle that the compounds present obey Beer's law of absorption, may be applied to other gas mixtures. These may be of C_6 hydrocarbons, i.e., pentanes and pentenes as well as many others.

Beer's law of absorption states that

$$I = I_0 e^{-\alpha l p}, \quad (1)$$

where I is the intensity of light of a particular wave-length transmitted by a sample of material, I_0 is the intensity of light of the same wave-length incident upon the sample, α is a constant which depends upon the material in the sample and the wave-length under investigation, l is the thickness of the sample, and p represents the concentration of the material in the sample which in case of gas would be the pressure. The constant α may assume different values depending upon what system of units is used for l and p .

For the successful application of the method of Brattain and others² to multicomponent mixtures, such that there is appreciable overlapping of the absorption bands of the different components, it is very important that each obey Beer's law of absorption. This is true for the mixtures named above as well as for others in both the gas and liquid phase. However, for some of the light gases such as CO , CO_2 , SO_2 , CH_4 , N_2O , HCl , etc., Beer's law is not obeyed and a somewhat different technique must be used. For these gases, the light absorption depends upon the total pressure in the absorption cell as well as the partial pressures of the absorbing materials. For example, a cell, containing 10 cm of Hg pressure of methane alone, will not absorb the same as if the cell had in addition to the same pressure of methane a partial pressure of some gas such as H_2 which does not absorb in the infra-red. More will be said about this later.

This anomalous pressure dependence of the absorption of the light gases has been known for

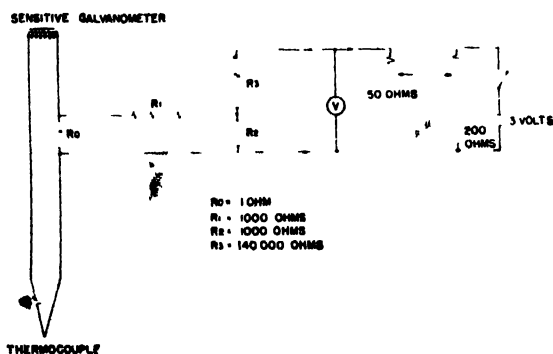


FIG. 1 Null circuit used to measure radiation intensities.

¹ Barnes, Gore, Liddel, and Williams, *Infra-Red Spectroscopy* (Reinhold Publishing Corporation, New York, 1944).

² Brattain, Rasmussen, and Cravath, *J. App. Phys.* 14, 418 (1943).

³ Norman Wright, *Ind. Eng. Chem. Anal. Ed.* 13, 1 (1941).

⁴ April-May, 1944 issue of *Transactions of the Faraday Society*, devoted to, "The Application of Infra-Red Spectra to Chemical Problems."

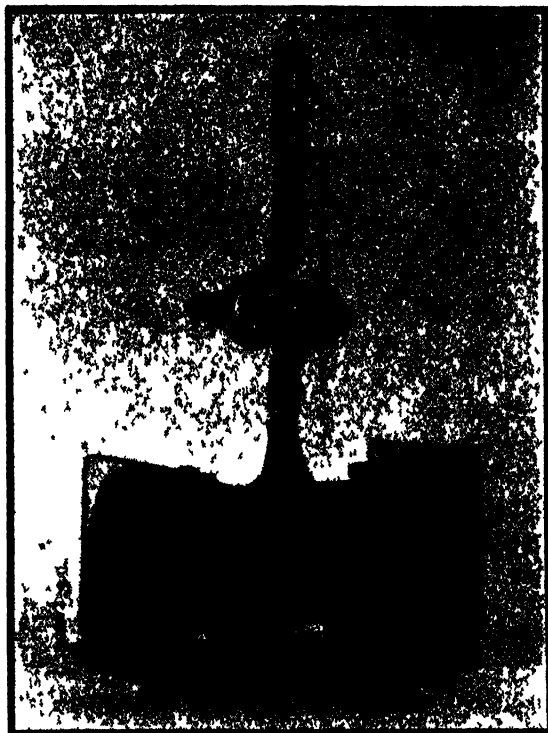


FIG. 2 Absorption cell assembly

some time and various investigators have studied the effect in different materials.⁵⁻⁹ D. M. Dennison has made a theoretical study of the factors affecting the shape and intensities of infra-red absorption lines.¹⁰ His results are based on the assumption that the principal factor in the broadening of an absorption line is the limitation of the length of wave train a molecule may absorb due to its perturbation by thermal collisions. He found that the absorption depends upon the number of molecules per unit volume, the temperature of the gas, and the effective molecular diameters. Experimental work was done by Cross and Daniels on the absorption of nitrous oxide as influenced by foreign gases. In line with Dennison's results they analyzed their data to obtain "optical" collision diameters for the foreign gases.

⁵ Eva. v. Bahr, Ann. d. Physik 29, 780 (1909); Physik. Zeits. 33, 585 (1910); Verh. d. D. Phys. Ges. 15, 710 (1913).

⁶ G. Hertz, Verh. d. D. Phys. Ges. 15, 673 (1913).

⁷ C. Schafer and F. Matossi, *Das Ultra-Rote Spectrum* (Verlagsbuchhandlung Julius Springer, Berlin, 1930), §30.

⁸ P. C. Cross and F. Daniels, J. Chem. Phys. 2, 6 (1934).

⁹ Other references can be found in article: Nielsen, Thornton, and Dale, Rev. Mod. Phys. 16, 307 (1944).

¹⁰ D. M. Dennison, Phys. Rev. 31, 503 (1928).

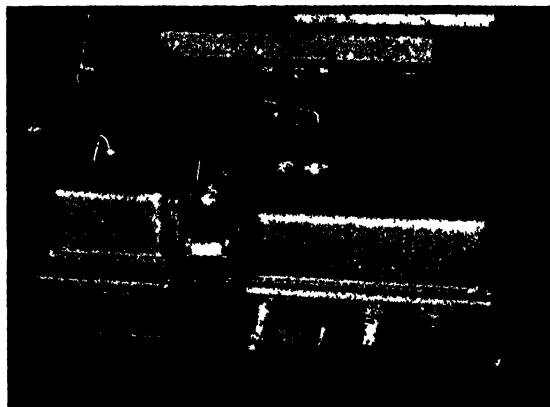


FIG. 3. Infra-red spectrometer used with added cell positioning equipment.

The data given below will illustrate the pressure broadening and how it is encountered in the making of actual analyses for some of the light gases. Several cases, for example, in the petroleum industry wherein it is desirable to make such analyses, are hydrogen rich gases such as are encountered in hydroforming, flue gas from a catalytic cracking unit, and burn-off gases from catalyst regeneration. Others working in the application of infra-red spectroscopy have doubtlessly encountered the same problems and perhaps evolved essentially the same solution. In that case the material of this article will illustrate the method and apparatus as used by us, the particular procedure, the accuracy obtainable by this procedure, and data of general interest with regards to pressure broadening.

INSTRUMENTAL DETAILS

The infra-red spectrometer used in this work was a Perkin-Elmer Model 12A. The thermo-

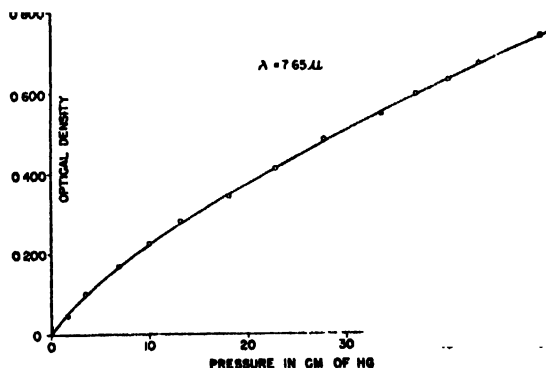


FIG. 4. Optical density vs. pressure for methane.

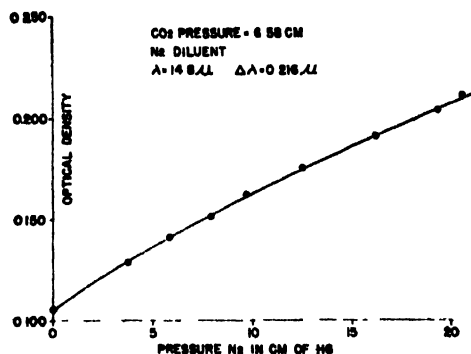


FIG. 5. Influence of N₂ on absorption of constant partial pressure of CO₂.

couple signals were fed into a Leeds and Northup, HS No. 2284-6, 0.05 $\mu\text{v}/\text{mm}$ sensitivity galvanometer. A null method was used rather than a direct photometric method wherein the galvanometer deflections are read. For this a controlled voltage was fed into the circuit containing the thermocouple and galvanometer in a direction opposite to the voltage signals due to radiation falling on the thermocouple. The controlled voltage was not measured directly but by measuring a large voltage to which it was in constant proportion. The circuit for the null method is to be seen in Fig. 1. Here the bucking voltage originates from the 3-volt battery which is connected directly across a potentiometer which is of 200-ohm resistance. The moving contact on the 200-ohm potentiometer provides a fairly good control but an additional fine control is available through the moving contact on the 50-ohm potentiometer. The resulting voltage tapped off by this variable resistance system is read on the voltmeter which is a Weston, Model 45, and which has 3 volts full scale as one of its ranges.

It is to be noticed that the circuit integral with and to the left of the voltmeter in Fig. 1 is continuous in the sense of no moving or variable contacts. Therefore, the voltage across R_0 is always a constant fraction of that across the voltmeter. If V represents the voltage across the voltmeter and V_0 the impressed voltage across R_0 , we may deduce for the case of balance by simple applications of Ohm's law

$$V_0 = \frac{1}{2}(R_0/R_2)V, \quad (2)$$

when R_1 and R_2 are equal and $R_2 \gg R_1$. With the resistances as indicated on Fig. 1 it is possible

to introduce a bucking voltage across R_0 up to 3.6 microvolts. In operation, when a signal to the thermocouple is to be determined, the mechanical zero for the HS galvanometer is first observed on a ground glass scale with a shutter in the light path and with the potentiometers adjusted to give zero voltage across the voltmeter. When the shutter is removed the potentiometers are adjusted to bring the galvanometer deflection back to the starting point. The voltage appearing on the voltmeter is then read and recorded as representative of the signal to the thermocouple. The system is very reproducible, successive readings giving values which have variations not larger than 0.2 percent. The 3-volt battery is in reality a combination of two Burgess 4F2H batteries connected in parallel for longer life. The batteries, resistances, and potentiometers are all mounted in an iron-walled case. This is important to protect the circuit from magnetic disturbances, which in a network of these impedances can cause serious disturbances.

The infra-red absorption cells used in this work have glass walls, an inside length of 9.5 cm, and have rocksalt windows. They are equipped with a stopcock, an inner part of a ground glass joint, and an exterior aluminum mounting frame. The bodies of the cells were painted black to minimize the amount of radiation of visible wave-lengths which may enter the monochromator chamber. This assembly is shown in Fig. 2. The aluminum frame has adjusting screws which allow the cell to be positioned as desired relative to it. Also the frame is equipped with slotted positioning brackets so that it may be clamped securely in place in the spectrometer. Two absorption cells

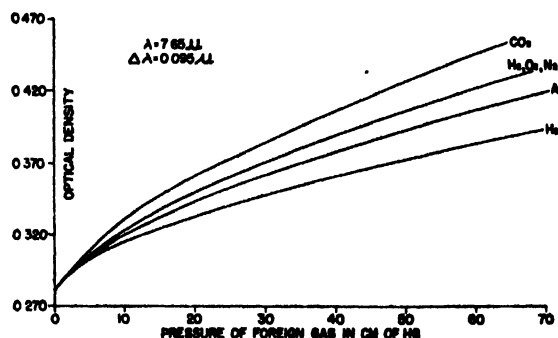


FIG. 6. Influence of gases: He, A, H₂, O₂, N₂, and CO₂ on absorption of constant partial pressure of CH₄.

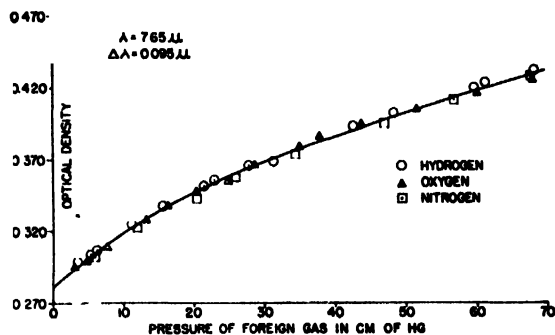


FIG. 7. Detail of influence of H_2 , O_2 , and N_2 on absorption of methane.

are used, one for the sample and one for comparison or incident energy readings. The two cells are mounted on a carriage mechanism which moves in such a manner that either cell may be moved into the optical path.

The instrument with the cell moving carriage with cells in place can be seen in Fig. 3. The carriage is equipped with a bar which the operator can use to position the cells. Adjustable stops insure that the cells always arrive in the same position in the light path. To remove the cells all that is necessary is to loosen two thumb-screws. The cells are evacuated and filled to measured pressures on a gas handling system separate from the instrument. Connected permanently to the thermocouple case in the instrument is a small mercury diffusion pump which is operated periodically to maintain a good vacuum in the presence of the radiation receiver. The apparatus here has been in use for some time for the analyses to be described as well as in the routine analysis of mixtures of the seven common C_4 hydrocarbons. For this, as well as for the present application, it has been very satisfactory and allows considerable speed of data taking. When it is desired to change the pressure in the absorption cell it is merely removed and engaged by means of its standard taper ground glass joint, into a gas handling system designed specifically for handling of routine samples.

In making blends and calibration mixtures for the analyses to be described as well as for other work of this type, it is important that the different components be thoroughly mixed. If the total pressure of the mixture in the absorption cell did not exceed 25 cm of Hg it was mixed in a separate, constant volume chamber.

This chamber was equipped with a thin steel diaphragm which is moved back and forth to mix the gases by means of exterior magnets. When the total pressure of the mixture in the cell exceeded 25 cm of Hg the mixing was done in the absorption cell itself. This was possible as the cell was equipped with a thin steel vane bent to fit the inside curvature of the cell. When the cell was in the operating position the vane lay flat against the bottom, out of the optical path. Mechanical motion of the cell such as shaking or rotating caused the vane to move back and forth, thoroughly mixing the gases.

DATA, CALIBRATION, AND PROCEDURE

For a gas obeying Beer's law a straight line curve is obtained when the observed optical density D ($D = \log I_0/I$) is plotted as a function of the pressure of the gas in the absorption cell. As stated above, some of the light gases do not obey Beer's law and data illustrating this may be seen in some of the following figures. In Fig. 4 is seen a graph of the optical density plotted against pressure for methane alone. As may be noticed, it is far from a straight line. It will be observed shortly that the general appearance of this plot is similar to those obtained when a constant pressure of methane is used and the partial pressure of a non-absorbing foreign gas is varied. In Fig. 5 may be seen a curve for the optical density for CO_2 plotted against partial pressure of added N_2 . It is observable that the effect of the N_2 is to increase the absorption very greatly. All the non-absorbing gases used were tested for absorption at the wave-lengths used. In each case they showed zero absorption.

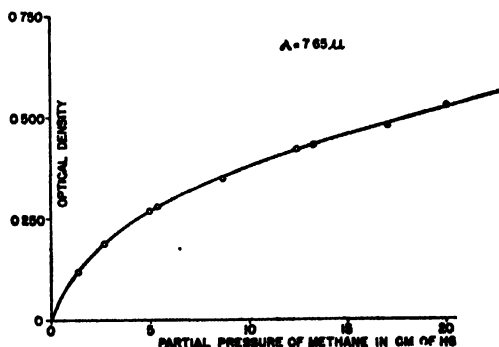


FIG. 8. Calibration curve constructed for analysis for methane.

Figure 6 shows the effect of various foreign gases on the absorption due to methane. The graphs shown in this figure were obtained by drawing smooth curves through the data obtained for each type of foreign gas. The amount of scattering of the data about the curve in each case is about the same as will be seen in Fig. 7. Here the ordinate represents the observed

TABLE I. Results for synthetic samples for methane analysis.

Sample No.	% CH ₄ (synthetic)	% CH ₄ (observed)	Δ %	Nature of remainder of sample
1	14.5	14.7	0.2	48% H ₂ ; 52% N ₂
2	7.6	7.7	0.1	N ₂
3	6.00	6.05	0.05	33% H ₂ ; 67% O ₂
4	22.5	22.3	0.2	O ₂
5	4.90	4.95	0.05	Air
6	3.71	3.76	0.05	H ₂
7	1.29	1.36	0.07	Air
8	0.83	0.85	0.02	3.9% Air; 96.1% H ₂

optical density and the abscissa represents the partial pressure of the foreign gas added. The data were all taken at the wave-length $\lambda = 7.65 \mu$ where methane has its strongest absorption. The slits were used at a setting which gave a spectral slit width of 0.095μ at this point. The partial pressure of methane was always constant, being 13.15 cm of Hg. The foreign gases added, namely, helium, argon, oxygen, nitrogen, hydrogen, and carbon dioxide, have no infra-red absorption at this wave-length and the large change in optical density produced is a manifestation of the pressure broadening effect mentioned earlier. It is to be noticed that the effects of the various gases are different, some of them being responsible for a greater increase in optical density than others. A study of these data in terms of the pressure broadening theory of Dennison and others has been made.¹¹ In Fig. 6 the effect due to nitrogen, hydrogen, or oxygen is shown as one curve; Fig. 7 shows in more detail the effects of these latter gases.

For the analyses in which we are interested here Fig. 7 is of primary importance. This shows that the effects of hydrogen, nitrogen, and oxygen as added gases are nearly the same. Thus it is possible to construct a single empirical calibration curve which will serve for the analysis of

¹¹ N. D. Coggeshall and E. L. Saier, Bull. Am. Phys. Soc. 21, 17 (1946).

methane in the presence of any one or mixture of the three. This is important as the off gases for which the method is used vary with regards to the relative concentrations of H₂, N₂, and O₂.

Figure 8 shows the calibration curve constructed for this analysis. For every determination the total pressure in the absorption cell was the same, i.e., 80 cm of Hg. The points were obtained by blending known amounts of methane and hydrogen together, admitting to the cell, adjusting the pressure to 80 cm, and determining the optical density. Figure 8, therefore, represents the measured optical density *versus* partial pressure of methane, the total cell pressure being constant.

With this curve the actual analysis of sample consists only of the steps of: admitting the sample into the cell to the pressure of 80 cm, determining the optical density, and from the curve reading off the partial pressure. Table I gives some results for a series of synthetic samples analyzed by means of the previously constructed calibration curve. The table shows the percent of methane as blended, the percent as determined by the analysis, the nature of the remainder of the sample, and the percent difference in terms of total sample between the synthetic and observed concentrations. The table not only illustrates the obtainable accuracy but also shows it to be independent of the nature of the remainder of the sample. Hundreds of samples have been analyzed for methane by this method and the results at all times have been satisfactory. The time required for a complete analysis, including introducing the sample, taking the data, and making the calculations, is about 15 minutes under routine conditions.

An analysis somewhat similar to the one just

TABLE II. Composition of blends used to construct calibration curves for flue gas analyses.

Blend No.	% CO ₂	% SO ₂	% CO	% Air
1	9.5	2.0	3.4	85.1
2	12.7	2.0	6.6	78.7
3	3.9	1.7	8.7	85.7
4	2.5	1.1	10.8	85.6
5	1.6	0.7	10.9	86.8
6	7.6	1.3	4.7	86.4
7	7.0	0.4	1.4	91.2
8	4.9	1.2	2.9	91.0
9	6.5	0.6	1.3	91.6

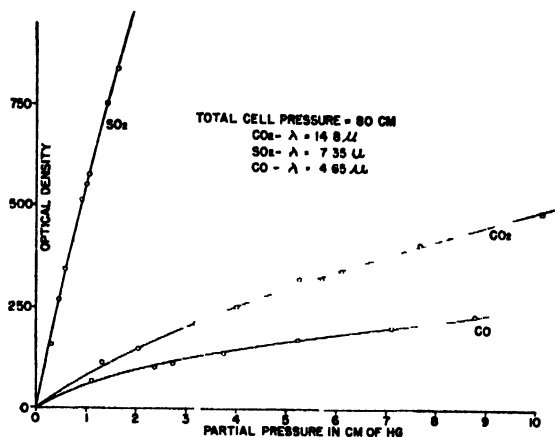


FIG. 9. Calibration curves for analysis of flue gas for CO_2 , SO_2 , and CO .

described, but involving more components, is the analysis of flue gases and burn-out gases from catalyst regeneration for CO_2 , CO , and SO_2 . In this case also the remainder of the samples varies, containing different amounts of O_2 , H_2 , and N_2 . Since for these samples the major part is generally a mixture of these three and since the concentration of any one of the three gases, CO , CO_2 , or SO_2 , is generally less than 10 percent it was decided to try the same method as used for the methane analysis.

A calibration curve for each of the three compounds CO , CO_2 , and SO_2 was constructed as was done for methane. In constructing these, blends of the three gases and of air were made and examined as before. The wave-length used for CO_2 was 14.8μ , for SO_2 it was 7.35μ , and for CO it was 4.65μ . Each gas has a maximum of absorption at about these wave-lengths and there is no overlapping between their absorption bands at these points. As before the blend was admitted to the absorption cell to the same pressure each time, this being 80 cm of Hg. In making the blends, care was taken to see that the concentrations of CO , CO_2 , and SO_2 were randomly distributed, that is to say: the concentration of any one of the three in any one of the blends was independent of the concentration of the other two. Table II gives the concentrations of the blends used for constructing the calibration curves.

The calibration curves are shown in Fig. 9. Here the uniformity and smoothness of the curves attests to the independence of the ab-

sorption of the three gases. In this figure the optical density observed at each wave-length was plotted against the partial pressure of the gas which absorbed at that point. The slits were set to a spectral slit width of 0.216μ for readings of 14.8μ ; 0.117μ for readings of 7.35μ ; and 0.085μ for readings of 4.65μ . Smaller slit widths could be used but the ones employed allowed the Globar element to be operated with about 150-watt supplied power. At this power level the life of each element is quite long. The shape of the curves indicates the deviations from Beer's law to be the same type as for methane.

Although the smoothness of the calibration curves is a manifestation of the internal consistency of the data, the usefulness and accuracy of the curves were tested by a further set of synthetic blends. The results are shown in Table III. Here it may be seen that the average error in terms of total sample is about 0.1 percent. The accuracy for CO is somewhat surprising in view of the flatness of the calibration curve for this compound.

As flue gas analyses may be used for process

TABLE III. Comparison between blended synthetic samples and concentrations measured by infra-red absorption for flue gas analyses.

Sample No	Compound	Synthetic	Observed	% Dev.
1	CO_2	10.5	10.4	0.1
	SO_2	2.1	2.3	0.2
	CO	4.2	3.9	0.3
2	CO_2	6.7	6.9	0.2
	SO_2	1.4	1.4	0.0
	CO	9.4	9.5	0.1
3	CO_2	6.2	6.2	0.0
	SO_2	1.9	1.9	0.0
	CO	4.0	4.1	0.1
4	CO_2	4.0	4.2	0.2
	SO_2	1.3	1.3	0.0
	CO	2.6	2.5	0.1
5	CO_2	2.7	2.8	0.1
	SO_2	1.0	1.0	0.0
	CO	5.2	4.8	0.4
6	CO_2	1.7	1.9	0.2
	SO_2	0.6	0.6	0.0
	CO	3.4	3.1	0.3
7	CO_2	7.0	7.1	0.1
	SO_2	1.9	1.9	0.0
	CO	6.1	6.3	0.2
8	CO_2	4.4	4.6	0.2
	SO_2	1.2	1.2	0.0
	CO	4.0	3.9	0.1

control it is important that the complete procedure be as simple as possible. This is important since it allows them to be done on night shifts by relatively untrained personnel. The equipment and method as described above have been used for process control analysis for a fluid catalytic cracking unit for some time and hundreds of samples have been handled for this purpose. One operator is sufficient to carry out all steps and the complete analysis, including calculations, can

be done on a routine basis in about 35 minutes or less. This includes duplicate determinations of the optical densities at each wave-length value.

The authors are indebted to Mr. N. F. Kerr for aid in developing some of the equipment used, to Dr. Morris Muskat for various suggestions in the application of infra-red absorption methods, and to Dr. Paul D. Foote, Executive Vice-President of Gulf Research and Development Company for permission to publish this material.

Methods of Minimizing Lead Loss in Emissivity and Resistivity Determinations

L. MALTER, *RCA Victor Division, Radio Corporation of America, Lancaster, Pennsylvania*

AND

D. B. LANGMUIR, ** Office of Scientific Research and Development, Washington, D. C.*

(Received January 12, 1946)

Two means are described for minimizing the effects of lead loss in determining the electrical properties of thin filaments as a function of temperature. In both cases the filaments are supported at the mid-points of filamentary U's. In one case the U's are made of finer wire than the main filament so that lead loss is reduced. In the other case, alternating current is passed through the U's so that the temperature of the main filament is maintained uniform down to its points of support.

DURING the course of work some time ago on the properties of tantalum filaments,¹ it was desired to extend the measurements to as low temperatures as possible. At these low temperatures difficulties were encountered because of the fact that the cooling effect of the filament supports extended up to and beyond the potential leads. Since in resistivity determinations it is essential that the filament temperature be uniform over the region (between the potential leads) whose electrical constants are being measured, it becomes necessary when following standard technique to use inordinately long filaments, unless they are of other than gossamer-like thickness.

To obviate this difficulty, recourse was had to one of the two expedients which are described below. In the first of these (see Fig. 1) the

filament, whose properties are to be determined, is welded at each end to the mid-point of an inverted U of considerably smaller size wire. By proper proportioning of the diameters of filament and U's, it is possible to secure uniform temperature of the filament not only over the region between the potential leads but in fact almost to the filament supports. A little experimental work may be necessary to arrive at the correct ratio of diameter of filament wire to U wire for securing satisfactory operation over the range of interest. In our work, a ratio of two to one was found satisfactory.

The second method illustrated in Fig. 2 is preferable in that no preliminary experimental work is required, and in that the filament temperature can be made uniform to its very ends for all temperatures. The inverted U's which support the filament in this case are made of the same size wire as the filament itself. With only the d.c. from the battery flowing through the U's to the filament, the U's run considerably cooler

* Formerly with RCA Laboratories, Princeton, New Jersey.

¹L. Malter and D. B. Langmuir, *Phys. Rev.* **55**, 743 (1939).

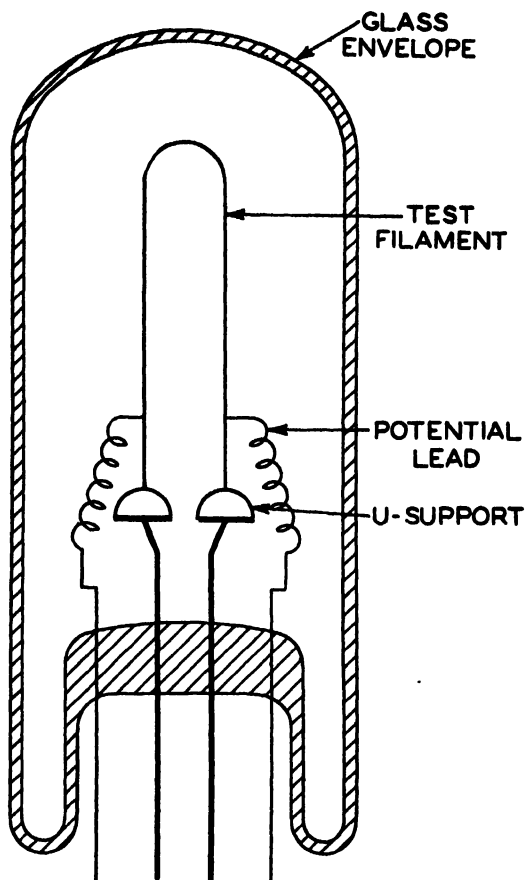


FIG. 1. Fine wire U supports for minimizing lead loss effects.

than the filament, thus resulting in a cooling of the latter to some distance from their points of support. However, by passing alternating current through the U's by means of external transformers, the temperature of the U's may be raised so that the main filament temperature is uniform right up to its points of support, the alternating current through the U's being controlled by means of small autotransformers as shown. Since no alternating current flows through the filament itself, the determination of the

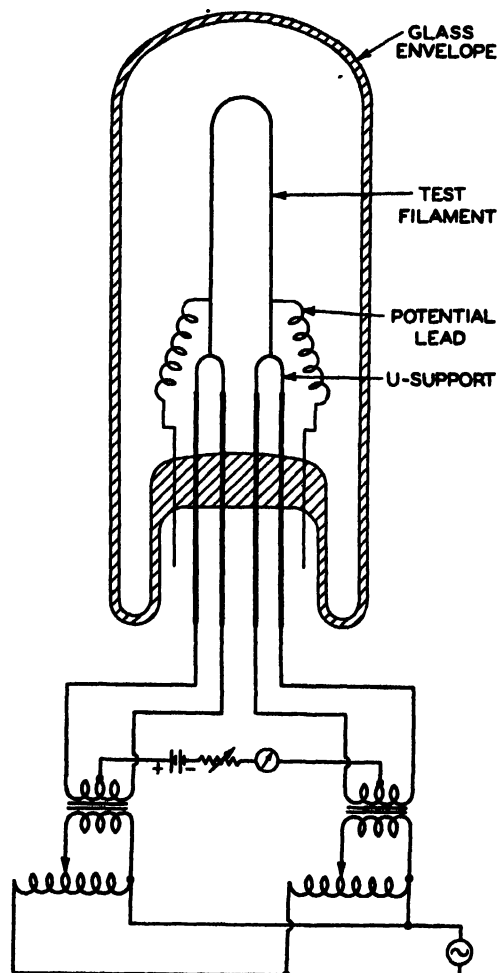


FIG. 2. Auxiliary heat method for minimizing lead loss effects.

electrical characteristics by d.c. methods between the potential leads is unaffected.

By use of these methods, it has been possible to extend the range of measurements to the limit of visibility. It appears that by a process of extrapolation, it should be possible to secure data beyond this limit.

Theory and Application of the Parallel Plate Plastometer*

G. J. DIENES AND H. F. KLEMM
Bakelite Corporation, New York, New York

A method has been established for the measurement of the viscosity of high polymers at low rates of shear in the range 10^4 to 10^9 poises using a parallel plate plastometer. This is based on a mathematical criterion for separating the viscous portion of the deformation from the "elastic" and "delayed elastic" components. Experimentally, the plate separation is measured at a given temperature as a function of time. The theory furnishes a relation, which is also the criterion for predominantly viscous deformation, between viscosity, plate separation, applied load, and time. This relation, a modified form of Stefan's equation, is used for calculating the viscosity from the experimentally observed quantities. The method has been applied to polyethylene and vinyl chloride-acetate resin compounds. The viscosity-

temperature behavior of these materials is shown to be simple over the temperature range studied; that is, log viscosity varies linearly with the reciprocal of the absolute temperature. Data are presented which show that polyethylene resins and polyethylene resin-paraffin wax mixtures follow Flory's relation; that is, log viscosity varies linearly with the square root of the weight average molecular weight. Accordingly, the parallel plate plastometer offers promising possibilities for the empirical determination of the weight average molecular weight of these materials. Data are also presented on plasticized vinyl chloride-acetate resin systems which point to a close parallel between the effects of increasing temperature and increasing plasticizer concentration.

INTRODUCTION

ONE of the most important properties of plastic materials is their ability to flow under the action of applied stresses at elevated temperatures. The quantitative evaluation of flow behavior in terms of applied stress and temperature is of great practical importance and theoretical interest.

It is well known that the behavior of plastic materials under the action of stresses is very complex because elastic and viscous deformations

take place simultaneously. The nature of stress-strain and deformation-time (constant load) curves has received the attention of many investigators in recent years.¹ It is generally agreed that a combination of elastic and viscous properties gives rise to the characteristic deformation-time or "creep" curves of high polymers.

A typical deformation-time curve is illustrated in Fig. 1. A quantitative separation of this curve into its elastic and viscous components has not yet been accomplished, although a

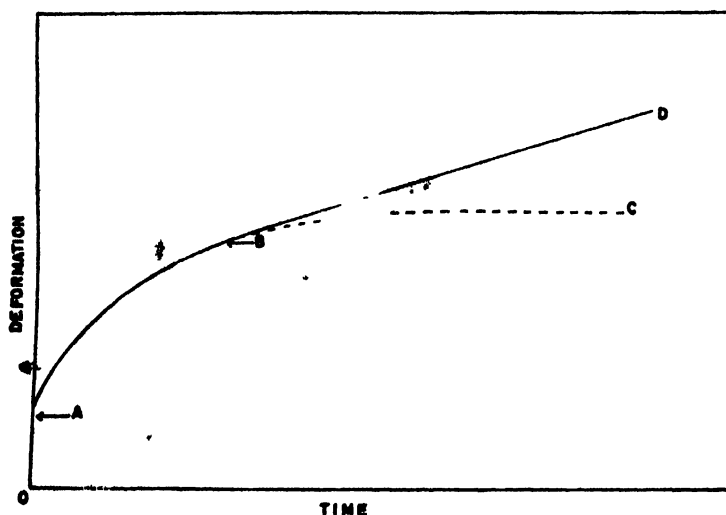


FIG. 1. Typical deformation-time curve.

* Presented at the Annual Meeting of the Society of Rheology, October 26 and 27, 1945.

¹ M. Reiner, *Ten Lectures on Theoretical Rheology* (Nordemann Publishing Company, New York, 1943). R. F. Tuckett, *Chem. Ind.* 430 (Nov. 1943). D. R. Morey, *Ind. Eng. Chem.* 37, 255 (1945). G. W. Scott Blair, *Survey of General and Applied Rheology* (Pitman Publishing Corporation, New York, 1944). T. Alfrey, *Quart. App. Math.* 3, 143 (1945); *J. Chem. Phys.* 12, 374 (1944).

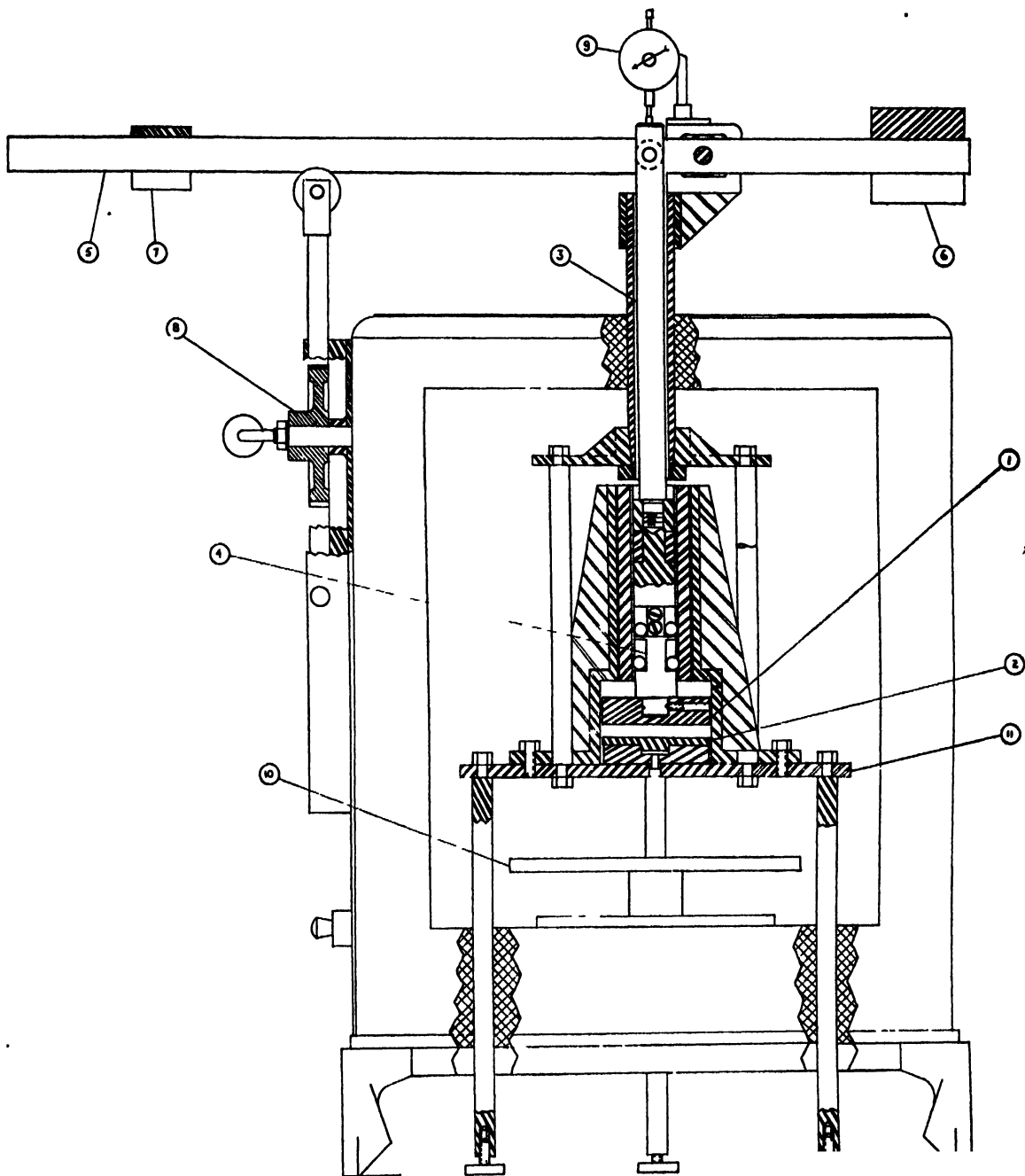


FIG. 2. Parallel plate plastometer.

qualitative separation into regions of predominance is possible. With reference to Fig. 1, this is done as follows:

(1) The first region is represented approximately by the portion of the curve OA . This represents a predominantly "elastic" deformation which takes place rapidly and is recovered similarly when the load is released.

(2) The second region is represented approximately by the portion of the curve AB , and is often referred to as the "delayed elastic" deformation. This deformation is characterized by a rapidly changing slope and may take place over a considerable period of time. It approaches a constant value as indicated by the broken line BC . The part BC can be observed only when no viscous deformation takes place.

(3) Region three is represented by a straight line, BD .

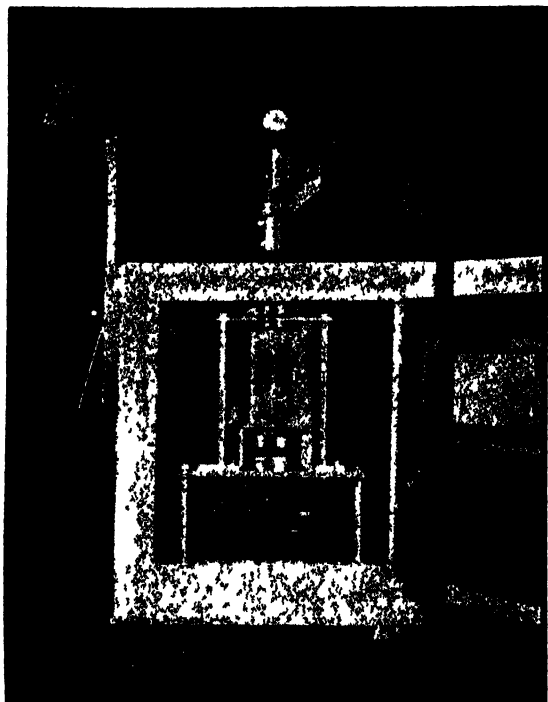


FIG. 3a. Plastometer showing 4-in. diameter plates separated.

It is generally agreed that this portion of the curve represents predominantly viscous deformation which is non-recoverable, and continues until the load is removed, the specimen is ruptured, or the test conditions are altered. This part of the curve is characterized by a constant slope. This does not apply, of course, at stresses and temperatures where a permanent deformation does not occur.

The three regions described above represent a rough separation of the deformation-time curve into its component parts. For a further understanding of the mechanical properties of plastics, it is essential to carry out this separation quantitatively. The final aim is to describe the complex mechanical behavior of high polymeric materials in terms of such fundamental constants as viscosity, modulus of elasticity, relaxation times, etc.

This paper describes an attempt to apply the parallel plate plastometer to the measurement of the predominantly viscous deformation (region three) free from "elastic" and "delayed elastic" effects.

MECHANICAL DETAILS

The improved parallel plate plastometer used for the evaluation of viscosity of plastics is shown in front elevation (sectioned for clarity) in Fig. 2. The instrument consists of two circular

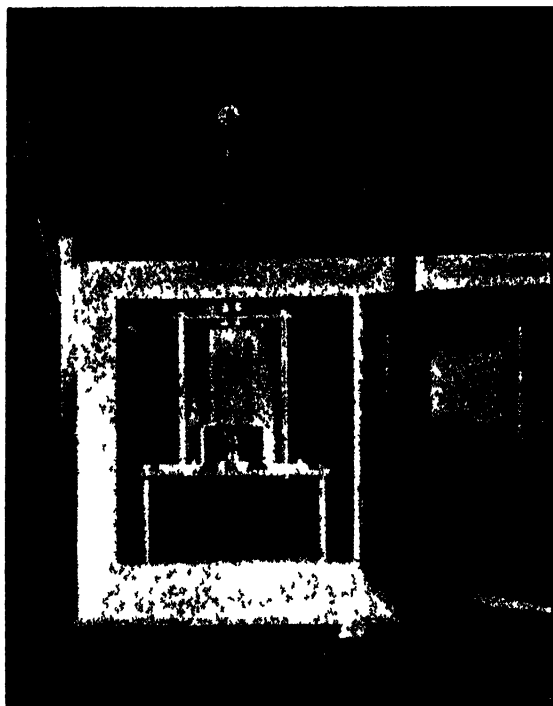


FIG. 3b. Deformation under load test (A.S.T.M. D 621-B).

parallel plates, both either 4.00 or 1.125 in. in diameter, the upper (see 1, Fig. 2) movable, the lower (2) stationary. The upper plate is attached to the vertical shaft (3) which is aligned and guided in the low friction ball-bearing assembly illustrated (4). The test load is applied by means of the beam (5), mounted at the upper end of the shaft (3), and the two movable weights (6 and 7). The dial indicator (9), which makes contact with the loading shaft, is fastened to the frame or reference system of the plastometer, and indicates the plate separation at any instant. A hand-operated rack and pinion mechanism (8) on the left-side of the oven permits the application of the test force without opening the oven door. The plastometer is supported by three legs, fitted with leveling screws, which project through the bottom of the oven.

Figures 3a and 3b are photographs of the plastometer which further illustrate its mechanical details. Figure 3a shows the apparatus with both the inner and outer oven doors open. The turntable (10, Fig. 2) underneath the plastometer proper, supports several clamps which hold the specimen assemblies together when preheating is done outside of the test area of the plates.

Figure 3b shows the equipment fitted with special plates (1.125 in. in diameter, one square inch in area) which are required in some deformation under load tests (A.S.T.M. D 621, Method B). These plates are also used when the plastometer is operated as Type 1 (see Theory), i.e., the plates have the same area as the ends of the specimen.

The important advantages of the parallel plate plastometer described above are listed below:

- (1) Low friction losses. All movable points of contact are provided with ball bearings to minimize friction.
- (2) Wide range of test loads (up to about 60 kilograms (132 lb.)).
- (3) External application of test load.
- (4) Dial indicator is attached to the plastometer frame and is located outside of the oven for accurate trouble-free operation at all test temperatures.
- (5) Small specimen door, located in inner of two oven doors, minimizes temperature changes due to sample insertion and removal.
- (6) Parallel plates removable for ease of cleaning and permitting change from one type of plates to another.

A more recent model of this instrument is equipped with a loading beam calibrated in units of 0.1 k.g.s.

EXPERIMENTAL PROCEDURE

Calibration of the Plastometer

The calibration of the plastometer was accomplished at room temperature by measuring the force transmitted by the upper plate to a platform balance. The applied force was varied by changing the position of the weights on the loading beam. The balance was placed in the oven beneath the main horizontal support member (11, Fig. 2) of the plastometer. This force was transmitted to the balance by a connecting rod which passed through the hole in the main horizontal member. Two threaded disks were attached to either end of the rod, and the assembly adjusted for length to obtain proper equilibrium position of the balance.

The calibration data in graphical form are shown in Fig. 4. It will be noted that both ranges are characterized by linear relations between load and the distance D_L . Furthermore, the observed calibration agreed very closely with that obtained by calculation.

Test Specimens

The test specimens were right circular cylinders, consisting of a stack of disks 0.63 in. (1.60 cm) in diameter, cut from sheet stock 0.040 to 0.080 in. (0.101 cm to 0.202 cm) thick. The specimen mass in grams was equal to the mass of an integral number of disks closest to twice the density of the material. Such a specimen has a height of about 0.394 in. (1.0 cm) and a volume of about two cm^3 .

Test Procedure and Calculations

An outline of the test procedure for determination of viscosity by the parallel plate plastometer is given below.

- (1) Adjust the oven to the desired test temperature, and allow sufficient time (at least one hour) to insure thermal equilibrium in the apparatus.
- (2) Position the loading weights along the lever arms to give the desired test load.
- (3) Set the dial gauge so that it indicates minus two divisions (one div. = 0.001 in.) when the plates are together under test load (no specimen). This corrects for the thickness of the two one-mil sheets of cellophane which are

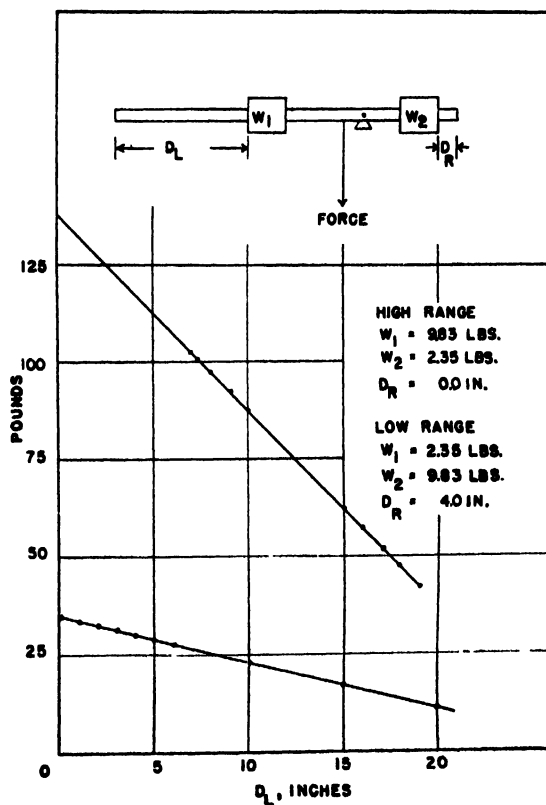


FIG. 4. Plastometer load system calibration.

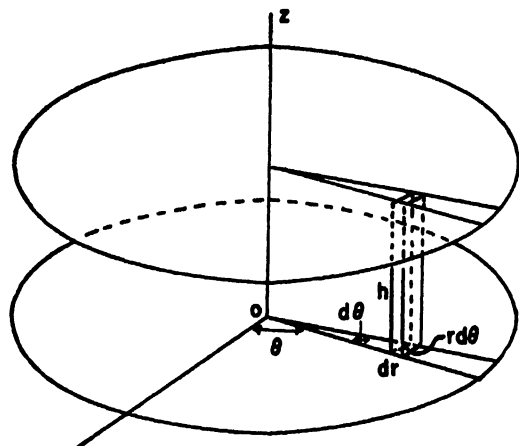


FIG. 5. Schematic diagram of plastometer illustrating coordinate system.

placed on the two ends of the specimen, and fixes the zero point for specimen height readings.

(4) Center the specimen (weighed to the nearest milligram) between two $4 \times 4 \times 0.001$ in. pieces of plain cellophane, and insert this assembly centrally between the parallel plates. If preheating of the specimen is required, this may be done either in the plastometer or in special clamps provided for this purpose.

(5) Apply the test load and read the separation, h , of the two parallel plates on the dial gauge to an accuracy of ± 0.0002 (0.2 of a division) as a function of time.

(6) Plot on linear coordinate paper, as the test proceeds, the reciprocal of the fourth power of the plate separation, $1/h^4$, expressed in cm^{-4} , as ordinate against the time in seconds. A table of $1/h^4$ values (in cm^{-4}), in the range $h = 0.010$ in. to $h = 0.150$ in., facilitates this operation.

(7) Continue the test until a straight line of slope, m , is defined in the range of h corresponding to $R:h = 10$ (R = radius of specimen) and $h = 10$ divisions (0.010 in.). Loads and temperatures must be adjusted to satisfy these requirements. Otherwise, the data are not simply interpretable.

(8) Calculate the viscosity, η , from the following equation:

$$\eta = 8.21 \times 10^4 (W/m^2), \quad (16)$$

where η = viscosity, poises, W = applied load, kilograms, m = slope of plot of $1/h^4$ vs. time, $\text{cm}^{-4} \text{ sec}^{-1}$, and V = volume, obtained from specimen mass and density, cm^3 .

THEORETICAL CONSIDERATIONS

Preliminaries

In any physical test certain measurements are made, such as length, time, temperature, etc. From these measurements, it is desired to determine quantities which are characteristic of the material. The purpose of theory is to express the physical law which relates the variables measured in the test with the properties of the material.

In the parallel plate plastometer tests, interest is focused on the flow characteristics of the material. Probably the most important constant which describes the flow characteristics of a material is its viscosity.

In the case of the parallel plate plastometer, it is desired to find an equation relating the plate separation, the time, and the viscosity. This problem, which is to find the equation of motion of two plates moving toward each other under the action of a force with a viscous medium between the plates, has been solved,² and the solution successfully applied to the parallel plate plastometer.

Derivation of Stefan's Equation from the Fundamental Equation of Viscous Flow

The general equation of motion of a Newtonian fluid of viscosity η (neglecting body forces such as gravity) is given by the vector equation

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \eta \nabla^2 \mathbf{v} + \frac{1}{3} \eta \nabla \nabla \cdot \mathbf{v}, \quad (1)$$

where ρ = density of the fluid, $\mathbf{v}$ = velocity vector, and p = fluid pressure. If the fluid is incompressible, the equation of continuity requires that

$$\nabla \cdot \mathbf{v} = 0.$$

If the velocity of the fluid elements is small, the term on the left in the square of $\mathbf{v}$ may be neglected. For slow motion of an incompressible viscous fluid, therefore³

$$\rho \partial \mathbf{v} / \partial t = -\nabla p + \eta \nabla^2 \mathbf{v}. \quad (2)$$

In cylindrical coordinates, Eq. (2) in scalar form becomes

$$\begin{aligned} -\frac{\partial p}{\partial r} + \eta \nabla^2 v_r &= \rho \frac{\partial v_r}{\partial t}, \\ -\frac{1}{r} \frac{\partial p}{\partial \theta} + \eta \nabla^2 v_\theta &= \rho \frac{\partial v_\theta}{\partial t}, \\ -\frac{\partial p}{\partial z} + \eta \nabla^2 v_z &= \rho \frac{\partial v_z}{\partial t}. \end{aligned} \quad (3)$$

Equations (3) are to be solved subject to the

² M. J. Stefan, Akad. Wiss. Wien. Math.-Natur. Klasse. Abt. 2., **69**, 713-735 (1874).

³ L. Page, *Introduction to Theoretical Physics* (D. Van Nostrand Company, New York, 1930), pp. 224-230.

boundary conditions defined by the parallel plate plastometer. In cylindrical coordinates, let the circular parallel plates be the planes (Fig. 5)

$$\begin{aligned} z &= 0; \\ z &= h. \end{aligned}$$

The mathematical theory is considerably simplified if the analysis is limited to that part of the data for which the plate separation, h , is small compared to the radius, R , of the specimen cylinder, the ends of which are in contact with the parallel plates.

Because of circular symmetry, v_θ is zero. As long as h is small compared to R , a good approximation is obtained by neglecting the velocity, v_z , normal to the plates, in comparison with the radial velocity, v_r . Thus,

$$\begin{aligned} v_\theta &= 0; \\ v_z &= 0. \end{aligned}$$

It will be assumed that there is no slippage at the plates, that is

$$v_r = 0 \text{ at } z = 0 \text{ and } z = h.$$

As a consequence, the radial velocity, v_r , changes rapidly over the short distance, h , from a finite value in the liquid to zero value on the plates. Under these conditions, then, $\partial v_r / \partial r$ is negligibly small compared to $\partial v_r / \partial z$. From symmetry, $\partial v_r / \partial \theta = 0$.

Steady-state flow is assumed, that is

$$\partial v_r / \partial t = 0.$$

Thus, the solution will not be applicable at the start of an experiment when transients due to $\partial v_r / \partial t \neq 0$ are not negligible.

With the above simplifications, Eq. (3) becomes

$$\frac{\partial p}{\partial r} = \eta \frac{\partial^2 v_r}{\partial z^2}, \quad (4)$$

where p is now a function of r only.

Experimental work on the parallel plate plastometer indicated that the above approximations are justified as long as $R > 10h$. Erratic viscosity measurements, which are difficult to interpret, are obtained when this condition is not obeyed.

Integrating Eq. (4) twice with respect to z and

inserting the boundary conditions,

$$v_r = 0 \text{ at } z = 0 \text{ and } z = h,$$

gives

$$v_r = -\frac{1}{2\eta} \frac{\partial p}{\partial r} (z-h)z. \quad (5)$$

Equation (5) shows that the velocity of the fluid in any plane, $\theta = \text{constant}$, varies parabolically in z . It is at this point that M. J. Stefan² starts his derivation with the assumption of a parabolic law.

Consider now a cylindrical element (see Fig. 5) extending from one plate to the other, and contained between radii r and $r+dr$ and between two radial planes parallel to z an angle $d\theta$ apart. The flow through a surface element $r d\theta dz$ is $r d\theta v_r dz$. Then U , the flow per unit arc length between the planes $z=0$ and $z=h$ (in the direction of increasing r), is

$$\begin{aligned} U &= \int_0^h v_r dz, \\ &= -\frac{1}{2\eta} \frac{\partial p}{\partial r} \int_0^h (z^2 - zh) dz, \\ &= -\frac{h^3}{12\eta} \frac{\partial p}{\partial r}. \end{aligned} \quad (6)$$

Suppose that the plate $z=0$ remains fixed and that the plate $z=h$ remains parallel to it but moves in the z direction with the velocity dh/dt (dh/dt positive in the positive z direction). Because of this motion, the cylindrical element, $drrd\theta h$, will change in volume at the rate

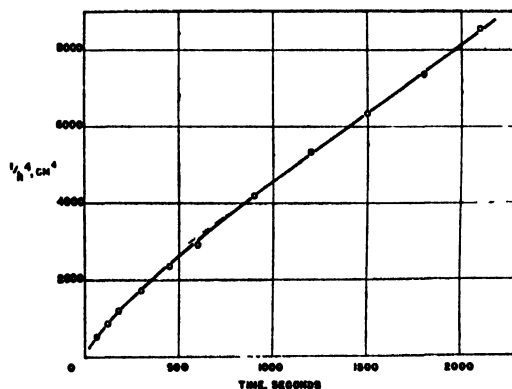


FIG. 6. Typical $1/h^4$ vs. time curve for polyethylene at 115°C.

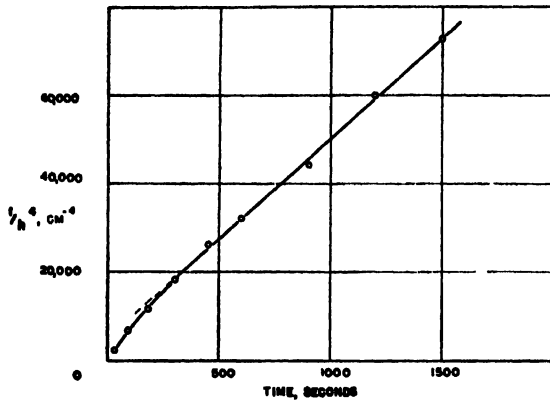


FIG. 7. Typical $1/h^4$ vs. time curve for polyethylene at 190°C .

$dr r d\theta dh/dt$. Evidently the difference between outflow and inflow must compensate for this change in volume, since the fluid was assumed to be incompressible. If the plate $z=h$ moves toward the plate $z=0$, then the rate of decrease of the volume element must equal the net rate of outward radial flow. Thus,

$$-dr r d\theta \frac{dh}{dt} = -\frac{\partial}{\partial r}(r d\theta U) dr. \quad (7)$$

Upon substitution of the expression for U and simplification

$$\frac{12\eta}{h^3} \frac{dh}{dt} r = -\frac{\partial}{\partial r} \left(r \frac{\partial p}{\partial r} \right). \quad (8)$$

Integration twice with respect to r gives

$$p = \frac{3\eta}{h^3} \frac{dh}{dt} r^2 + C' \ln r + C. \quad (9)$$

As p must be finite for $r=0$, C' vanishes. C is evaluated from the condition that at the outer boundary of the specimen, $r=R$, the pressure must be in equilibrium with the constant pressure, Π (usually atmospheric). Thus,

$$C = \Pi - \frac{3\eta}{h^3} \frac{dh}{dt} R^2,$$

and Eq. (9) becomes

$$p = -\frac{3\eta}{h^3} \frac{dh}{dt} (R^2 - r^2) + \Pi. \quad (10)$$

The total force acting on the plate in the positive z direction is

$$\int_0^R p 2\pi r dr.$$

The total force acting on the plate in the opposite direction is

$$F + \int_0^R \Pi 2\pi r dr,$$

where F = applied force (dynes). Therefore,

$$F + \int_0^R \Pi 2\pi r dr = - \int_0^R \frac{3\eta}{h^3} \frac{dh}{dt} \times (R^2 - r^2) 2\pi r dr + \int_0^R \Pi 2\pi r dr$$

or

$$F = -2\pi \frac{dh}{dt} \frac{3\eta}{h^3} \int_0^R (R^2 - r^2) r dr. \quad (11)$$

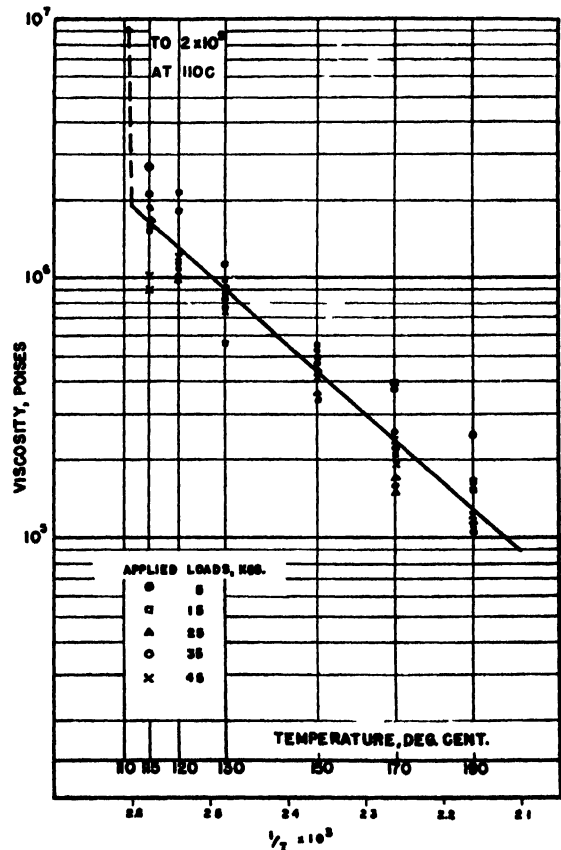


FIG. 8. Viscosity-temperature relation for polyethylene.

In the evaluation of this integral, two cases are treated.

Case 1: The viscous liquid completely fills the space between the plates so that $R=a$ at all times, where a is the radius of the plates. In this case, Eq. (11) on integration becomes:⁴

$$F = -\frac{3\pi\eta a^4}{2h^3} \frac{dh}{dt} \quad (12)$$

Equation (12) integrated with respect to t gives

$$\frac{1}{h^2} = \frac{4F}{3\pi\eta a^4} t + C_1, \quad (13)$$

where

C_1 = constant of integration.

Equations (12) and (13) are known as Stefan's equations.²

Case 2: The viscous material does not completely fill the space between the plates. Thus, the radial distance, R (radius of the specimen), changes with time, due to the motion of the upper plate. In this case, Eq. (11) is integrated from $r=0$ to $r=R$ where R now is a function of the time, t . This can be done by assuming that the material between the plates is always essentially in the form of a cylindrical sample. It has already been assumed that the viscous material tested is incompressible so that the volume, V , is a constant. In this case, integration of Eq. (11) leads to

$$F = -\frac{3\eta V^2}{2\pi h^5} \frac{dh}{dt} \quad (14)$$

Equation (14), on integration with respect to t , gives

$$\frac{1}{h^4} = \frac{8\pi F}{3\eta V^2} t + C_2, \quad (15)$$

where

C_2 = constant of integration.

The values of the constants of integration, C_1 of Eq. (13) and C_2 of Eq. (15), respectively, are the values of $1/h^2$ and $1/h^4$ at $t=0$, subject to the assumption that Stefan's equation is obeyed at

⁴ *The Mechanical Properties of Fluids*. A collective work. Applied Physics Series. (D. Van Nostrand Company, New York, 1924), pp. 113-119.

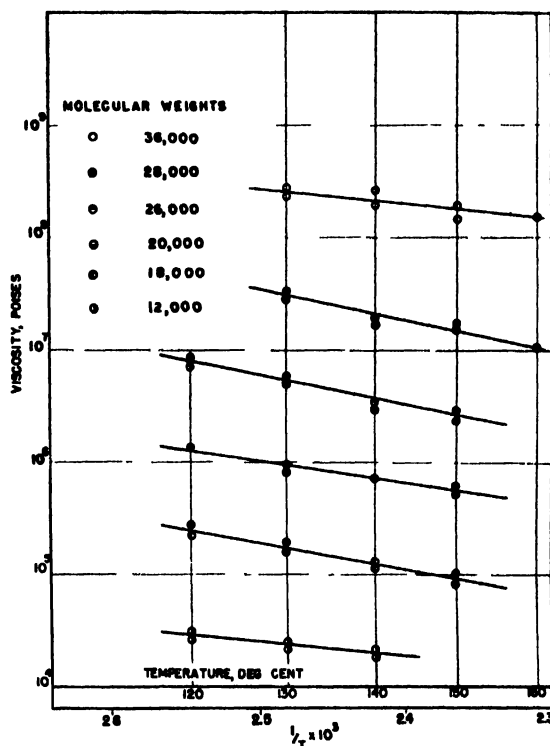


FIG. 9. Viscosity-temperature characteristics of various molecular weight grades of polyethylene.

$t=0$, and therefore do not correspond to the actual value of h at $t=0$. They are of no importance in the determination of the viscosity, η , from Eqs. (13) and (15), since η is calculable from the slope of the $1/h^2$ or $1/h^4$ vs. time curve.

Equations (13) and (15) describe the flow between two parallel plates moving toward each other in terms of the plate separation, h , the time, t , and the viscosity, η .

Calculation of Viscosity from Plate Separation and Time Measurements

In any parallel plate plastometer, a sample is compressed between the two plates, and the plate separation, h , is determined as a function of time. Two types of parallel plate plastometers may be considered corresponding to the two equations, Eqs. (13) and (15).

Type 1: The plates have the same area as the ends of the specimen. In this case, the space between the plates is completely filled with the viscous material at all times, and Eq. (13) is directly applicable in its integrated form. According to this equation, a plot of $1/h^2$ versus t will

be linear with the slope of the straight line depending inversely on the viscosity. An example of the use of this type will be given later.

Type 2: The plates are large compared to the area of the ends of the specimen. In this case, the radius of the contact area, R , varies with the plate separation, h , and Eq. (15) is applicable. According to this equation, a plot of $1/h^4$ versus t will be linear with the slope of the straight line depending inversely on the viscosity.

The parallel plate plastometer in this laboratory may be used as either Type 1 or 2. As a matter of experimental convenience, it has been operated almost exclusively as Type 2, and Eq. (15) has been used to calculate the viscosity. For this case, the viscosity may be expressed from Eq. (15) in terms of the applied load; slope, m , of the $1/h^4$ vs. time curve; and specimen volume as follows:

$$\eta = 8.21(10)^8 (W/mV^2). \quad (16)$$

In this equation, the applied force, F , previously expressed in dynes, was replaced by W expressed in kilograms, m is expressed in $\text{cm}^{-4} \text{sec}^{-1}$, V in cm^3 , and η in poises.

If the experimental data obtained on a material follow the $1/h^4$ versus t straight line law, then the viscosity of the material can be calculated from Eq. (16). Accordingly, the linearity of the $1/h^4$ versus t plot was adopted in this work as the criterion for the measurement of predominantly viscous deformation. Other workers in this field, namely V. E. Meharg, L. E. Welch, and W. A. Miller;⁵ R. L. Peek;⁶ and N. M. Foote,⁷ have used the differential form as given by Eq. (14) for the calculation of viscosities. R. W. Griffiths,⁸ in a paper on the plasticity of rubber, considered the use of Eq. (15) for the calculation of viscosity (a simplified theoretical treatment is given by Healey in the appendix), and showed a $1/h^4$ vs. time curve for rubber. The possibility of using the linearity of the $1/h^4$ vs. time plot for separating the viscous deformation from other effects was apparently not investigated further.

It must be pointed out that this linear relation

⁵ V. E. Meharg, L. E. Welch, and W. A. Miller, unpublished data.

⁶ R. L. Peek, Jr., *J. Rheology* 3, 53, 345 (1932).

⁷ N. M. Foote, *Ind. Eng. Chem.* 36, 244 (1944).

⁸ R. W. Griffiths, *Trans. I. R. I.* 1, 308 (1925-26).

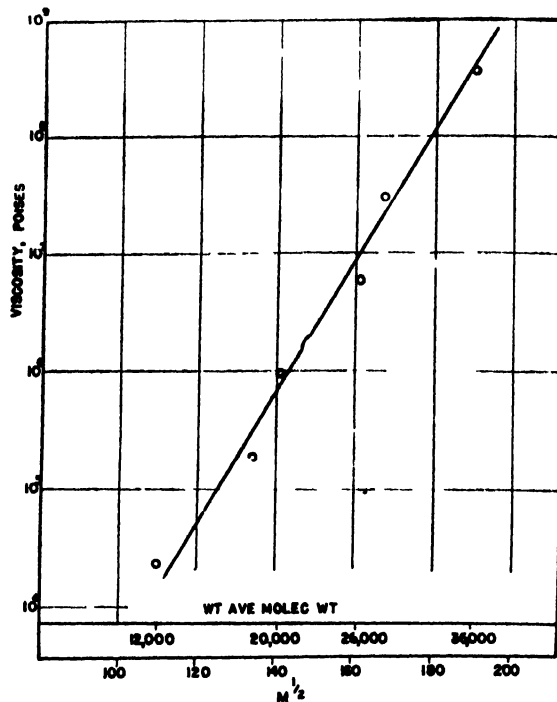


FIG. 10. Viscosity of polyethylene resins as a function of the square root of the weight average molecular weight. Temperature = 130°C.

was derived for Newtonian viscous flow. Thus the fact that a given material obeys this linear relation means that the material behaves as a Newtonian fluid at the rates of shear employed in the test. Mathematically, Dillon and Johnston⁹ were able to obtain approximate expressions for the average rate of shear in different types of plastometers. They showed that the average rate of shear for parallel plate plastometers is in the neighborhood of 0.1 sec^{-1} (extrusion plastometers may be run at rates of shear in the neighborhood of 500–1000 sec^{-1}). The parallel plate plastometer, therefore, measures viscosity at very low rates of shear. Non-Newtonian viscous flow is characterized by non-linearity of the rate of shear versus shear stress curve. These curves, however, may appear to be approximately linear for some non-Newtonian materials over the relatively small range and at the low rates of shear available in this instrument. Thus, it is impossible to determine, even by the tenfold variation of load available for the plastometer in this laboratory, the shape of the rate of shear

⁹ J. H. Dillon and N. Johnston, *Physics* 4, 225 (1933).

vs. shearing stress curve at other than very low rates of shear. Hence, the Newtonian or non-Newtonian character of the material remains indeterminate in parallel plate plastometer measurements. This instrument, combined with the mathematical technique developed, measures the absolute viscosity of very viscous high polymers in the range 10^4 to 10^9 poises, subject, of course, to the limitation noted above.

EXPERIMENTAL RESULTS ON SOME THERMOPLASTICS

Polyethylene

The viscosity of a polyethylene resin compound identified as DE-3401 Natural (polyethylene resin stabilized with 0.2 percent Akroflex C) was measured over an 80-degree temperature range at different applied loads.

Typical $1/h^4$ *vs.* t curves are shown in Figs. 6 and 7 at 115°C and 190°C, respectively. These curves were obtained after a thirty-minute preheat so that temperature equilibrium was established prior to load application. These graphs show that the $1/h^4$ *vs.* t curves are not linear until 300 to 1000 seconds have elapsed, with this period of non-linearity decreasing with increasing temperature. During this initial period of non-linearity, the $1/h^4$ *vs.* t curves are convex upward.

In order to explain the initial non-linear portions of these curves, two effects have to be considered. In the first place, the theory is not applicable in the very early stages of a test, since the experimental conditions do not satisfy the mathematical boundary conditions assumed in the derivation of Eqs. (13) and (15). However, after the first few seconds, non-linearity cannot be ascribed to this situation, but is due to the fact that the deformation is not simply viscous in character. Appreciable "delayed elastic" effects may be present which mask the viscous deformation. From the character of the three regions of a deformation-time curve, it is seen that the viscous deformation which is non-recoverable continues at a constant rate after the sum of the "elastic" and "delayed elastic" deformation has become constant (or at least negligibly small). Thus, it is expected that the criterion of viscous flow, namely the linearity of a $1/h^4$ *vs.* t plot,

will be obeyed except in the initial period when the "elastic" deformations are predominant. This is the experimental observation as represented by the curves of Figs. 6 and 7. After an initial period of non-linearity, the $1/h^4$ *vs.* t curves become straight lines, and these latter represent viscous flow. The decrease in time of the initial portion of non-linearity with increasing temperature is also in agreement with what is known about the temperature dependence of "delayed elasticity."¹⁰

The viscosity of the material is calculated from the slope of the linear portion of the $1/h^4$ *vs.* t curves by the use of Eq. (16).

The experimental results will be presented in the form of $\log \eta$ *vs.* $1/T$ curves. This plot will give a straight line if these materials follow the well-known viscosity-temperature law,

$$\eta = Ae^{(B/T)}, \quad (17)$$

where A and B = material constants, η = viscosity in poises, and T = absolute temperature. Activation energies for viscous flow, which are proportional to the constant B , may be compared if the slopes of the $\log \eta$ *vs.* $1/T$ curves are determined.

The experimental results on polyethylene (DE-3401) are shown in Fig. 8. This figure contains all the points of duplicate tests at five different loads (5, 15, 25, 35, and 45 kilograms) at temperatures of 110°, 115°, 120°, 130°, 150°, 170°, and 190°C. From the erratic spread of the experimental points at different loads, it is concluded that, although a change in load undoubtedly has some effect on the viscosity, no definite trend dependent on load was determined.

At 110°C the viscosity of polyethylene is about a thousand times as high as at 115°C, with the only reliable experiment at 110°C obtained under a 5 kg load (at higher loads the specimen was partially crushed). This result is in agreement with the known behavior of polyethylene which shows a phase transition¹¹ in the vicinity of this temperature.

The $\log \eta$ *vs.* $1/T$ plot is a straight line from 115°C to 190°C. It is to be noted that the variation of viscosity with temperature for poly-

¹⁰ R. F. Tuckett, *Trans. Faraday Soc.* **39**, 158 (1943).

¹¹ R. F. Clash, Jr. and L. M. Rynkiewicz, *Ind. Eng. Chem.* **36**, 279 (1944).

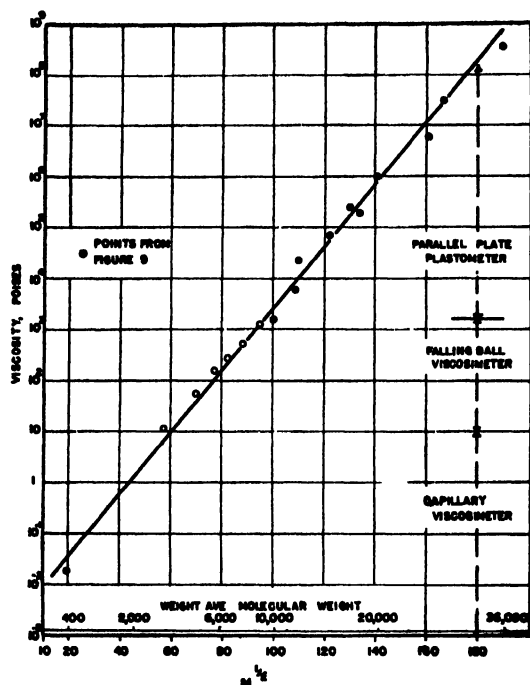


FIG. 11. Viscosity of the polyethylene resins and paraffin wax-polyethylene resin mixtures as a function of the square root of the weight average molecular weight. Temperature = 130°C.

ethylene is relatively small, roughly a tenfold change taking place over a range of 80 degrees, indicating a low activation energy for viscous flow.

The effect of preheating the specimen was investigated, and it was found that preheating had no appreciable effect on the measured viscosity. If the sample is not preheated, the initial portion of non-linearity of the $1/h^4$ vs. t curve may be somewhat longer, but the slope of the straight line portion is not affected. Accordingly, all further experiments were run without preheat as a matter of convenience.

The viscosity of polyethylene compounds (polyethylene resin stabilized with 0.2 percent Akroflex C) of various weight average molecular weights was also determined. Six polyethylenes of weight average molecular weight in the range 12,000 to 36,000 were available.

The weight average molecular weight (relative) of these materials was obtained by the Staudinger method as modified by Kemp and Peters.¹² The determination, made at 130°C in tetralin, was

¹² A. R. Kemp and H. Peters, Ind. Eng. Chem. **35**, 1108-1112 (1943).

based on the equation

$$M = \frac{K_{em} \log_{10} \eta_r}{C}, \quad (18)$$

where

M = weight average molecular weight,

η_r = relative viscosity,

C = base molar concentration of polymer in tetralin at temperature of determination (130°C),

= 4.0 g per liter of solution at 130°C/14, and

$K_{em} = 4.03 \times 10^4$ g per liter of solution at 130°C.

The experimental values of M used in this paper were taken from data obtained by Mr. C. S. Myers.¹³

The viscosity was measured for each material over a convenient temperature range. The results are given in the form of $\log \eta$ vs. $1/T$ curves in Fig. 9. Linear behavior is observed for all of these materials over the temperature range investigated. All of the curves have about the same slope, indicating that the activation energy for viscous flow does not change appreciably over this range of average molecular weights.

These data offer a test of the applicability of Flory's relation¹⁴ to polyethylene. The parallel plate viscosity for these resins may be identified with the "melt viscosity." It was shown by Flory that the melt viscosity of polymerized linear polyesters varies with the weight average molecular weight and the temperature according to the equation

$$\log \eta = D + \frac{B}{T} + CM^1, \quad (19)$$

where η = melt viscosity, T = absolute temperature, M = weight average molecular weight (relative), and D , B , and C = constants.

TABLE I.

Weight average molecular weight	Approximate time (seconds) to establish linearity of $1/h^4$ vs. t at 130°C
12,000	120
18,000	150
20,000	350
26,000	400
28,000	700
36,000	1000

¹³ C. S. Myers, unpublished data.

¹⁴ P. J. Flory, J. Am. Chem. Soc. **62**, 1057 (1940).

Figure 10 is a plot of $\log \eta$ versus the square root of the weight average molecular weight at 130°C. A straight line is obtained confirming Flory's relation for polyethylene.

Since the slopes of the $\log \eta$ vs. $1/T$ curves are essentially independent of M (Fig. 9), the constants (D , B , and C) of Eq. (19) can be determined. This gives

$$\log \eta = -6.14 + \frac{2015}{T} + 0.05M^{\frac{1}{2}}, \quad (20)$$

which was found to agree approximately with all the experimental results.

From these relations, it is evident that the parallel plate viscosity can be used with advantage in estimating weight average molecular weights. The use of the parallel plate plastometer extends considerably the range of viscosity that may be measured conveniently, thereby expanding the range of molecular weights that can be estimated.

It is interesting to consider the variation of "delayed elasticity" as a function of the average molecular weight. The length of time necessary to reach the linear portion of the $1/h^4$ vs. t curve may be considered an index of "delayed elasticity." The length of this time interval was estimated, at 130°C, from the $1/h^4$ vs. t curves. The data given in Table I were obtained.

TABLE II.

Compound B-660	
Vinyl chloride-acetate copolymer, grade VYNS	82.20
Vinyl chloride-acetate copolymer, grade VYHH	15.00
Lubricant, stabilizer, etc.	2.80
	100.00
Compound B-661	
Vinyl chloride-acetate copolymer, grade VYHH	97.35
Lubricant, stabilizer, etc.	2.65
	100.00

These data show that "delayed elasticity" increases markedly with increasing average molecular weight. It is to be noted that the measurements on these materials were made without any preheat. Since it requires about two minutes to obtain temperature equilibrium in the specimen, the two low molecular weight compounds show a "delayed elasticity" of very short duration at

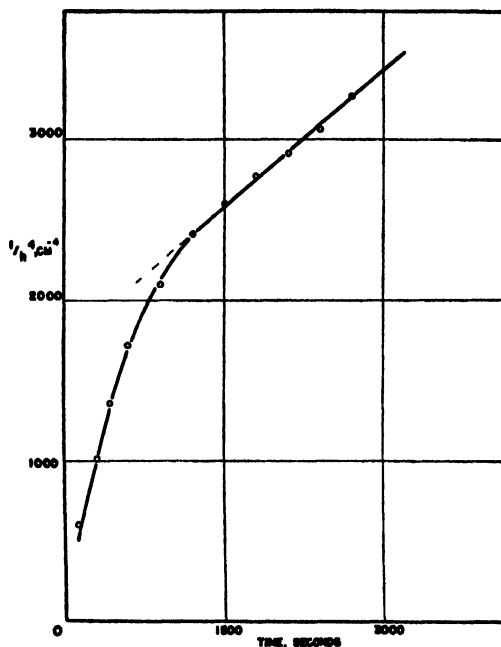


FIG. 12. Typical $1/h^4$ vs. time curve for a vinyl chloride-acetate resin compound (B 660) at 155°C.

this temperature. The trend with increasing weight average molecular weight, however, is unquestionable.

Paraffin Wax—Polyethylene Mixtures

Parallel plate viscosity measurements were carried out on a set of paraffin wax-polyethylene resin mixtures in the 50 percent to 100 percent resin content range. These compounds were stabilized mixtures of a refined low melting, household grade of paraffin wax ("Esso Wax," m.p. 52°C) and a standard polyethylene resin (Grade DYNH, wt. av. mol. wt. 20,000) containing 0.1 to 0.2 percent antioxidant.

Viscosity figures for the mixtures of 0 percent to 50 percent resin content were made available by Mr. C. S. Myers.¹³ The viscosity of the paraffin wax was determined by a Ubbelohde Viscosimeter (Type 1A); the viscosities of the mixtures (up to 50 percent resin content) were obtained by a falling ball viscosimeter. Weight average molecular weights, obtained by the method of Kemp and Peters as already described, were taken from Mr. Myers' data.

Application of Flory's equation (Eq. (19)) to these viscosity data at 130°C showed a remark-

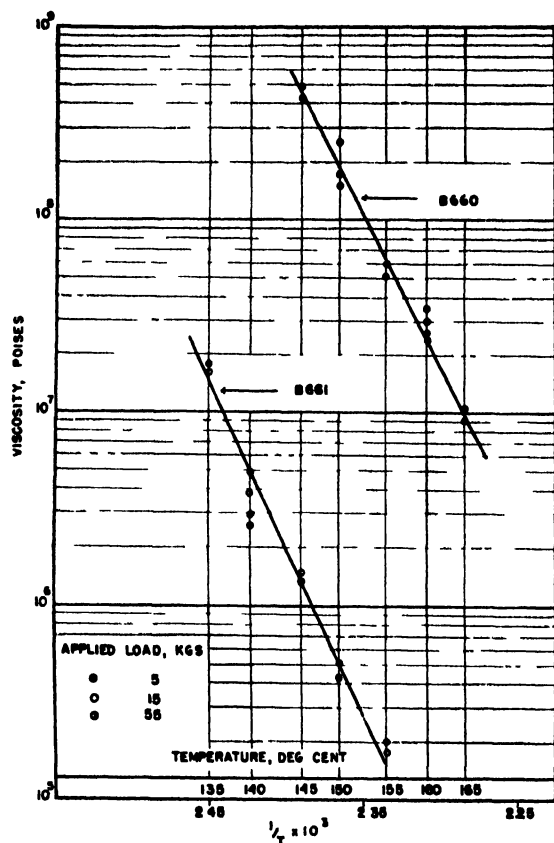


FIG. 13. Viscosity-temperature characteristics for two vinyl chloride-acetate resin compounds.

able degree of conformity. Figure 11 is a plot of $\log \eta$ versus the square root of the weight average molecular weight at 130°C. This figure contains all the data for the mixtures, paraffin wax alone, and all the polyethylenes of various molecular weights from Fig. 10. It will be noted that very good agreement is obtained among three different methods of viscosity measurement with the viscosity varying from 10^{-2} about 10^9 poises. It is also observed that Flory's relation evidently applies over a wide range of weight average molecular weights (400 to 36,000).

Vinyl Chloride-Acetate Resin Compounds

The theory and the discussion of the results for polyethylene apply with minor modifications to vinyl chloride-acetate resin compounds. Viscosity measurements were made on two unplasticized vinyl chloride-acetate resin compounds (composition in percent by weight) (see Table II).

A typical $1/h^4$ vs. time curve for compound

B-660 (at 155°C) is shown in Fig. 12. This curve is very similar to the curves for polyethylene, except that the "delayed elastic" region is much more pronounced. The straight line portion is not reached for about 25 minutes. Similar behavior was observed with compound B-661. These $1/h^4$ vs. t curves must be interpreted with caution due to these long initial periods of non-linearity. For the two compounds above, very good straight lines were obtained over the temperature range investigated. Viscosities were calculated from the slope of the straight line portions.

The viscosities follow the usual $\log \eta$ vs. $1/T$ straight line relation, as illustrated in Fig. 13. This figure also shows that the effect of changing the applied load is negligible. It is to be noted that the viscosity variation with temperature is much higher than for polyethylene, indicating a much higher activation energy for viscous flow.

This technique of viscosity measurement has

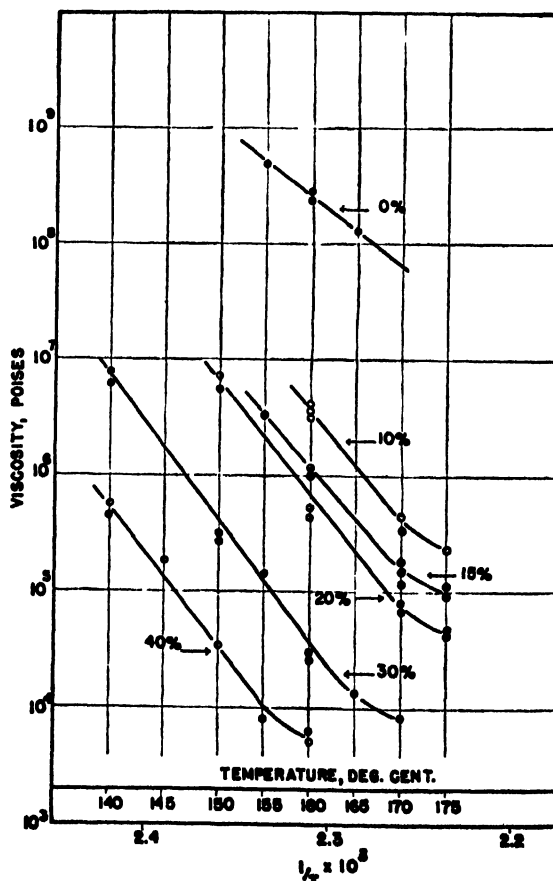


FIG. 14. Viscosity-temperature characteristics of VVNS+diethyl phthalate systems.

also been applied successfully to plasticized vinyl chloride-acetate resin compounds. The general formula used for these compounds was (percent by weight):

Vinyl chloride-acetate resin VYNS	97-X
Diocetyl phthalate	X
Stabilizers and lubricants	3
	100

The $\log \eta$ vs. $1/T$ curves at different plasticizer concentrations are shown in Fig. 14. It will be noted that, for the plasticized compounds, straight lines are obtained up to a certain temperature, above which the viscosity does not decrease as rapidly with increasing temperature as predicted by the straight line. It is also noteworthy that the slope of the $\log \eta$ vs. $1/T$ curves, that is, the activation energy for viscous flow, is essentially unchanged by the addition of plasticizer.

Figure 15 is a plot of $\log \eta$ vs. percent diocetyl phthalate at 160°C . This figure shows that very large changes in viscosity can be accomplished by varying the plasticizer content. The relation is linear, except in the 0 to 10 percent plasticizer range, pointing to a close parallel between the effects of increasing temperature and increasing plasticizer concentration.

The parallel plate plastometer has also been operated as Type 1; that is, the plates have the same area as the ends of the specimen. In this case, the sample has to be accurately lined up with the plates. For this test a $1/h^2$ vs. time straight line law is applicable (Eq. (13)). Experiments on polyethylene and vinyl chloride-acetate resins showed that this measurement of viscosity agrees with that obtained with the Type 2 plastometer. Experimentally, it has been more convenient to operate the plastometer as Type 2 because the alignment of the specimen is less critical.

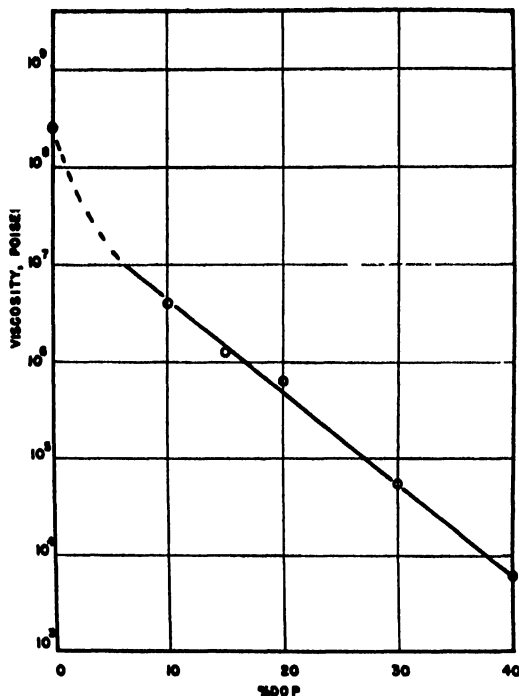


FIG. 15. Viscosity-plasticizer content curve for VYNS+diocetyl phthalate at 160°C .

ACKNOWLEDGMENT

The parallel plate plastometer used in this work was designed by Messrs. H. A. Swallow and W. R. Yaroshuk, and constructed by Tinius-Olsen Testing Machine Company. Grateful acknowledgment is also made to our associates in the Physics Division of the Bakelite Development Laboratories, particularly to R. F. Clash, Jr. and A. P. Wangsgard, for many helpful discussions on the various phases of this work. It is also a pleasure to acknowledge the helpful assistance derived from the earlier unpublished work on the parallel plate plastometer of Messrs. V. E. Meharg, L. E. Welch, and W. A. Miller of our Research Laboratories.

The General Relations for Flow in Solids and Their Application to the Plastic Behavior of Tire Cords

W. JAMES LYONS*

*Southern Regional Research Laboratory,** New Orleans, Louisiana*

(Received January 30, 1946)

Numerous researches on metals, rocks, glass, rubber, and textiles are cited to show that what may be regarded as *normal* creep in amorphous and polycrystalline solids conforms to the general relation

$$\epsilon = \epsilon_1 + at + b \log t,$$

where ϵ is the total strain at time t , ϵ_1 is a parameter interpretable as the approximate initial strain, and a , b are other parameters. A special form of this equation, having $a=0$, has been frequently applied. Reference is made to several other studies on the same class of solids, which have established for normal relaxation the relation

$$\sigma = \sigma_1 - \beta \log t,$$

where σ is the stress at time t , σ_1 is the stress at unit time, and β is a parameter. A theoretical foundation for both equations is provided by the reaction-rate theory of plastic flow. Observations in the present study indicate that creep extension in cotton and rayon tire cords over prolonged periods of time follows the above creep equation, and, in general, is not adequately represented by an equation omitting the term at . From the presence of this term in the equation, the existence of a component of the viscous type in tire cord growth is deduced. The pattern of creep recovery in cotton tire cord appears to be set by the behavior of the cord in creep extension

THE phenomena of creep and relaxation arising from flow in solids has been studied in a wide variety of materials ranging from soft bread doughs to the rocks in the crust of the earth. Attention has frequently been called to the technical importance of these flow phenomena in rubber products, textiles (particularly the industrial textiles, such as tire cord), high polymeric plastics, metals, and numerous other materials normally accepted as "solids." Considering the great diversity of chemical composition, molecular character, and general physical behavior, it is indeed remarkable that the laws describing the general pattern of creep and relaxation in a large number of these substances take the same two analytical forms. The only characteristic which is evidently common to all of this broad group of substances is a structure which is predominantly either amorphous or polycrystalline. This observation suggests that the types of behavior in creep and relaxation common among these substances are to be associated with randomness of structure.

* Present address: The Firestone Tire & Rubber Company, Akron 17, Ohio.

** One of the laboratories of the Bureau of Agricultural and Industrial Chemistry, Agricultural Research Administration, U. S. Department of Agriculture.

Any attempt to bring experimental data on a wide variety of otherwise unrelated materials into conformity with a general law should proceed with caution. On the other hand, the scientific ideal of systematized knowledge dictates that such possible generalizations be recognized, though they should not become the means for the entrenchment of misinformation. In the field of plastic-solid flow, a dispersive trend has prevailed, new empirical equations having been advanced with no apparent attempt having been made to correlate the data with earlier, applicable equations. The benefit of a *general*, empirical law, especially if there is a concomitant explanatory theory, lies in its suggestion of the elementary processes giving rise to exceptional behavior in particular cases.

THE LAW OF CREEP

The deformation resulting from creep under constant load at constant temperature in a large number of materials is in conformity with the relation

$$\epsilon = \epsilon_1 + at + b \log t, \quad (1a)$$

where ϵ is the strain at time t , ϵ_1 is a parameter having the dimensions of strain, and a , b are

parameters depending on the load, material, and ambient experimental conditions. $\epsilon(t)$ in Eq. (1a) is to be regarded as a function describing the creep process; in general, some other function of time describes the deformation in the initial stage. Thus, while the two processes are continuous physically, Eq. (1a) is invalid at $t=0$, and no attempt to evaluate $\epsilon(t)$ at $t=0$ should be made. The parameter ϵ_1 may be interpreted as an *extrapolated* initial strain, for if unity on the time scale used with Eq. (1a) is taken as the instant at which the initial strain is completed, and the insignificant contribution of at during the very brief initial stage is neglected, it will be seen that the deformation at later times behaves as if its initial value were ϵ_1 . In general, ϵ_1 only approximates the strain actually occurring during the initial stage. Equation (1a) was advanced by Weaver¹ to represent the total creep extension of various steels at constant stress and temperature in tests lasting, in some cases, five years. The parameter corresponding to ϵ_1 was negative, however, and was not given the interpretation outlined above.

An approximate relation of a type similar to Eq. (1a) omits the term at , thus implicitly placing $a=0$, and giving

$$\epsilon = \epsilon_1 + b \log t. \quad (1b)$$

The creep component of the total deformation is here represented solely by a logarithmic function of the time. The possibility of so representing creep appears to have been first recognized by Boltzmann² in 1874. He used an equation of the type of (1b) to express the torsional deformation in a filament as a function of time.

In order to define the function $\epsilon(t)$ so that it will be continuous at $t=0$, the argument of the logarithmic term may be written in the binomial form $(ct+1)$. Applying this modification, Eqs. (1a) and (1b) become

$$\epsilon = \epsilon_1 + at + b \log (ct+1), \quad (1c)$$

and

$$\epsilon = \epsilon_1 + b \log (ct+1), \quad (1d)$$

where c is a parameter having the dimension of inverse time. For moderately large values of ct ,

however, the relation

$$b \log (ct+1) \approx b \log c + b \log t$$

holds very closely. Hence, if ϵ_1 in Eqs. (1a) and (1b) is assumed to include the constant term $b \log c$, no distinction between Eqs. (1a) and (1c) or (1b) and (1d) can be made on the basis of conformity with experiments in which $ct \gg 1$. For this reason, the form $\log (ct+1)$ appears to be more comprehensive than is usually necessary for the representation of creep. Equation (1c), which may be regarded as the most general law of creep under constant load, has not been found in the literature. Equation (1d), however, has been used, as indicated below.

Since its application to creep in steel by Weaver,¹ Eq. (1a) has been employed to represent the slow yield of limestone, and of a single crystal of halite under compression,³ as well as the creep in molded phenolic plastics.⁴ Other applications of this equation have been cited by Findley.⁵ Furthermore, there are in the literature several researches on creep which have yielded results evidently in conformity with Eq. (1a); the reports, however, do not cite the equation. Among these may be mentioned the tensile tests of Trouton and Rankine⁶ and Andrade⁷ on lead, those of Bailey⁸ on molybdenum steel at elevated temperatures, and the three-hour compressive tests of Lyons⁹ on copper and lead. The data published on creep under tension in eight organic plastics of various types¹⁰⁻¹² appear to be in accord with Eq. (1a) when moderate deforming stresses are employed. Examination of the results of Findley^{5,12} indicates that under high stresses the creep in cellulose acetate proceeds as if a in Eq. (1a) had a negative value, which is a rather unacceptable condition. Griggs³ has noted a

¹ D. Griggs, J. Geol. 47, 225 (1939).

² D. Telfair, T. S. Carswell, and H. K. Nason, Mod. Plastics 21, 137 (Feb. 1944).

³ W. N. Findley, Symposium on Plastics (Philadelphia, A.S.T.M., 1944), p. 118.

⁴ F. T. Trouton and A. O. Rankine, Phil. Mag. [6] 8, 538 (1904).

⁵ E. N. Andrade, Proc. Roy. Soc. A84, 1 (1910); A90, 329 (1914).

⁶ R. W. Bailey, J. App. Mech. (Trans. A.S.M.E.) 58, (A)1 (1936).

⁷ W. J. Lyons, J. App. Phys. 9, 641 (1938).

⁸ B. Chasman, Mod. Plastics 21, 145 (Feb. 1944).

⁹ J. Delmonte and W. Dewar, Mod. Plastics 19, 73 (Oct. 1941).

¹⁰ W. N. Findley, Mod. Plastics 19, 71 (Aug. 1942).

¹ S. H. Weaver, Trans. A.S.M.E. 58, 745 (1936).

² L. Boltzmann, Pogg. Ann. 7, 624 (1876), Sitz. K. Akad. Wiss. Wien, Math.-Naturwiss. Classe 70, 275 (1874).

similar behavior in creep in shale. The physical interpretation of Eq. (1a), however, provides an explanation of this behavior at high stress.

In the field of textiles, several studies have yielded results which are amenable to expression by means of Eq. (1a). Among these are the long duration creep results of Leaderman¹³ on silk, and viscose and acetate rayon filaments. The data on single Nylon filaments, however, present an anomaly with reference to Eq. (1a) which resembles that found in cellulose acetate under high stress. The creep results of Busse and co-workers¹⁴ on rayon tire cord suggest that both the parameters a and b increase with temperature. In the study of Dillon and Prettyman,¹⁵ creep in cotton and rayon cords, at three different temperatures, is represented by straight lines on an *elongation vs. logarithmic-time* graph, indicating a value of zero for a . However, the curves found by Dillon and Prettyman for Nylon cord show a definite upturn at later times. In the experimental portion of the present study, results on tire cord creep extending into thousands of minutes were obtained. For these results Eq. (1a), as will be shown, provides a good representation.

The creep of soft copper¹⁶ under constant tensile load, as well as that of viscose yarns,¹⁷ has been shown to be in accord with Eq. (1d), the coefficients b and c being functions of certain material characteristics and experimental conditions. The omission of the constant rate term at from the equation, as made in Eqs. (1b) and (1d), appears to be applicable only in short duration tests, in which the viscous type of flow represented by at has little opportunity to become appreciable. Equation (1b) has been used to represent tensile creep in rubber, glass, and annealed copper, platinum, silver, and lead wires.¹⁸ It has also been applied to creep results on cotton and rayon fibers in tension¹⁹ and lead cylinders in

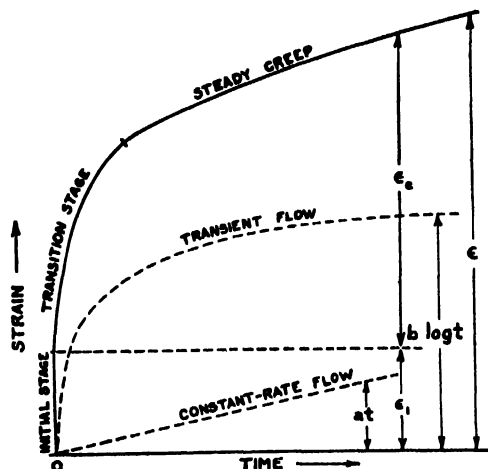


FIG. 1. Hypothetical diagram of the components of total creep deformation, as applied to terms in Eq. (1a).

compression²⁰ in thirty-minute tests. That creep in soft iron, also, is not a direct function of the time, but a function of the logarithm of the time, has been noted.²¹

THE LAW OF RELAXATION

The relaxation of stress at constant deformation also has been studied in a variety of materials. The general, empirical relation, for behavior at constant temperature, may be written

$$\sigma = \sigma_1 - \beta \log t, \quad (2a)$$

where σ is the stress at time t , σ_1 is a parameter interpretable as the initial stress, with the same limitations as were put on the interpretation of ϵ_1 , and β is a parameter depending on the initial stress, the material, and the particular temperature of the experiment. A variant of Eq. (2a) has been announced, and we give it forthwith:

$$\sigma = \sigma_1 - \beta \log (\gamma t + 1), \quad (2b)$$

where γ is a parameter like c .

Equation (2b), which, like Eq. (1c), is continuous at $t=0$, was given in 1904 by Trouton and Rankine⁶ for the decay of longitudinal stress in lead wire. In the same article they indicate that Eq. (2a) fits the data on the relaxation of torsional stress in a lead wire. In subsequent research, Eq. (2b) was applied to stress relaxation under constant strain in concentrated gels of

¹³ H. Leaderman, *Elastic and Creep Properties of Filamentous Materials and Other High Polymers* (Washington, Textile Foundation, 1943).

¹⁴ Busse, Lessig, Loughborough, and Larrick, *J. App. Phys.* **13**, 715 (1942).

¹⁵ J. H. Dillon and I. B. Prettyman, *J. App. Phys.* **16**, 159 (1945).

¹⁶ A. Nadai and E. A. Davis, *J. App. Mech. (Trans. A.S.M.E.)* **58**, (A)7 (1936).

¹⁷ J. J. Press, *J. App. Phys.* **14**, 224 (1943).

¹⁸ P. Phillips, *Proc. Phys. Soc. London* **19**, 491 (1905).

¹⁹ R. L. Steinberger, *Textile Research* **6**, 325 (1936).

²⁰ W. J. Lyons, *J. App. Phys.* **10**, 651 (1939).

²¹ H. Cassebaum, *Ann. d. Physik* **34**, 106 (1911).

gelatine,²² and recently, it has been applied to results on rayon yarn.¹⁷ It has been observed that Eq. (2a) represents the decay of stress in rubber under tension.¹⁸ The graphical plots of relaxing stresses in rubber against $\log t$ yield straight lines.²³ Equation (2a) has also been found to hold for relaxation in cotton, rayon, and Nylon tire cords at various temperatures.²⁴

No equation for relaxation containing a viscous-flow term such as at in Eq. (1a) has been found in the literature. There are some data, such as those of Trumpler²⁵ on relaxation at constant elongation in steel, which seem to call for such a term (with a negative sign), but in such cases the major portion of the relaxation is satisfactorily represented by Eq. (2a).

THEORETICAL CONSIDERATIONS

At least four equations^{5, 7, 26, 27} which have been advanced to relate stress σ , strain ϵ , and time t during flow in various solids involve power functions of t combined with coefficients having incomprehensible dimensions. In general, the power is not integral. In many of the applications, with a suitable choice of coefficients and exponents, good fits to experimental data have been made. The utility of the equations, however, is limited to convenient analytical expression, for purposes of extrapolation or prediction. Even aside from the difficulty of interpreting fractional powers of t , these empirical equations provide no insight into the physical processes occurring during plastic deformation. Equations (1a-d) and (2a, b), on the other hand, are not only interpretable in terms of gross physical effects, but find explanation in a theory of molecular mechanism.

As long ago as 1904 it was recognized by Trouton and Rankine⁶ that the plastic deformation of solids involves a component of the elastic type and one of the viscous type. The former consists of flow which gradually decreases with

time, and is more or less recoverable on removal of the deforming load. It has been designated by a seemingly endless number of names, among which "transient flow," used here, is one. The relation of these components (represented by dashed curves) to the total deformation ϵ (the solid curve) during creep, is shown in Fig. 1. Such a graphical analysis may be made without reference to a specific equation. Figure 1, however, lends itself well to the identification of the terms in Eq. (1a). The term ϵ_1 represents the deformation occurring in the initial stage, though as indicated above, numerically the representation is merely approximate. In the second term, a is the constant rate of flow characterizing the viscous component, which persists indefinitely. The transient flow is expressed with good fidelity by the final term $b \log t$. As Eq. (1a) indicates, the total strain at any instant during steady creep is given by the simple arithmetic sum of the components. The strain due to creep alone is $\epsilon_c = at + b \log t$.

The processes outlined in Fig. 1 have been interpreted in terms of another well-known creep equation,²⁸ namely,

$$\epsilon = \epsilon_1 + \lambda t + \epsilon_\infty [1 - \exp(-t/\tau)], \quad (3)$$

where λ is the constant rate of viscous flow, ϵ_∞ is the final, maximum strain due to transient flow, and τ is the Maxwellian relaxation time for the system. This equation was developed for a system of models of the type considered by Maxwell in the derivation of his classical equation for relaxation. Involving the evaluation of four coefficients, instead of three as does Eq. (1a), Eq. (3) is less manageable. Furthermore, as was found in the present study on tire cord, the fit to experimental data which is finally achieved is not as good as that which Eq. (1a) readily gives. Equations of the type of (3), however, have found favor, presumably because of their theoretical foundations, which until recently Eqs. (1) and (2) have lacked. Except insofar as the transient term $b \log t$ is an approximation to $\epsilon_\infty(1 - \exp(-t/\tau))$, Eqs. (1) and (2) are not derivable from Maxwell's equation. Integration of the latter equation for

²² A. O. Rankine, *Phil. Mag.* [6] **11**, 447 (1906).

²³ A. Tobolsky and H. Eyring, *J. Chem. Phys.* **11**, 125 (1943).

²⁴ H. R. Wakeham, Southern Reg. Res. Lab., private communication.

²⁵ W. E. Trumpler, Jr., *J. App. Phys.* **12**, 248 (1941).

²⁶ A. A. Michelson, *Proc. Nat. Acad.* **3**, 319 (1917); **6**, 122 (1920).

²⁷ P. G. Nutting, *J. Frank. Inst.* **191**, 679 (1921); **235**, 513 (1943).

²⁸ J. M. Burgers, *First Report on Viscosity and Plasticity* (Committee of Academy of Sciences at Amsterdam, Amsterdam, 1939), second edition, p. 28.

constant deformation leads to

$$\sigma = \sigma_1 \exp(-t/\tau), \quad (4)$$

as the expression for the relaxation of stress. This equation, however, has been notoriously poor in representing experimental data. While it may be in accord with data on a few materials, it fails for all those cases in which the corresponding empirical Eq. (2a) holds. Trouton and Rankine⁶ were among the earliest to abandon Eq. (4) in favor of Eqs. (2a) or (2b). Since the Maxwell equation is empirically unsatisfactory in the simplest applications, there appears to be little reason to accredit the underlying theory, and attempt to interpret more complicated processes in terms of it, as use of Eq. (3) implicitly does.

It is on the transient flow that most theoretical interest in creep and relaxation has centered. It will be noted that it is in the term for this component that Eqs. (1a) and (3) differ. The transient flow reappears in relaxation processes and is represented in Eqs. (2a) and (2b) by $\beta \log t$, or $\beta \log(\gamma t + 1)$. Some of the earliest analytical considerations of solid flow^{7,18} associated the transient part with local rupture, rotation, or reorientation of the submicroscopic elements of the material. Transient flow, as the integration of localized movements, was distinguished from the shearing process in continuing viscous flow. The later ideas considered the transient flow to be imposed on the viscous. This view is in line with the diagram of Fig. 1.

While qualitative theory has retained this picture of the general character of transient flow, the justification for the empirical, logarithmic form has been lacking. It is possible to set down a general differential equation for the process of deformation in solids, from which the initial and transient terms in Eqs. (1) and (2) can be derived and correlated. The treatment, however, is a mathematical formalism, not unlike those accompanying application of Maxwell's equation; it contributes nothing toward establishing the logarithmic form as the unique expression for specific molecular processes. The descriptive, reaction-rate theory of Eyring²⁰ does provide a basis for the logarithmic form in a mechanism at the molecular level. This theory leads to the

following relation between stress and time:

$$\tanh \frac{B\sigma}{2} = \tanh \frac{B\sigma_1}{2} \exp(-ABGt), \quad (5)$$

where σ_1 is the initial stress, G is an elastic modulus, and A and B are parameters involving temperature among other factors. Tobolsky and Eyring²⁰ have shown that, applied to longitudinal creep under constant load, beyond the range in which the cross section is changing rapidly, Eq. (5) leads approximately to Eq. (1b). For stress relaxation at constant elongation they derive Eq. (2a).

The constant-rate term at in Eq. (1a) does not emerge from the reduction of Eq. (5), which refers to the integration of localized slipping processes, such as might occur in crystalline solids or high polymeric network-structures. These processes consist of recrystallization arising from the assumption of new equilibrium positions by displaced atoms, and the reformation of ruptured bonds between chain molecules. In the complete Eyring theory of solid flow, however, there appears an equation for the motion of the amorphous portion of linear-polymeric substances ("segment motion"). If this equation is integrated independently for constant stress, the resultant expression yields a strain which increases at a constant time-rate. The theory in its present form, however, does not employ the outright addition of this strain to that given by Eq. (5). It does not appear feasible to dismiss the constant-rate term at as an insignificant experimental deviation from the logarithmic law, for in some materials, as lead in creep-extension⁶ measurements for instance, the deformation resulting from viscous-like flow is the most prominent component. It may be concluded that in the creep of the numerous solids which behave in accordance with Eq. (1a), there is an actual underlying constant-rate flow corresponding to the term at . The constant-rate, viscous-type flow may be interpreted as a slow general distorting motion within a connected, reticular system, or between such systems, through a matrix of unassociated, amorphous substance. This slow movement could conceivably take place whether the motive force were an external load, or internal restoring stresses in the networks.

²⁰ H. Eyring, *J. Chem. Phys.* **4**, 283 (1936).

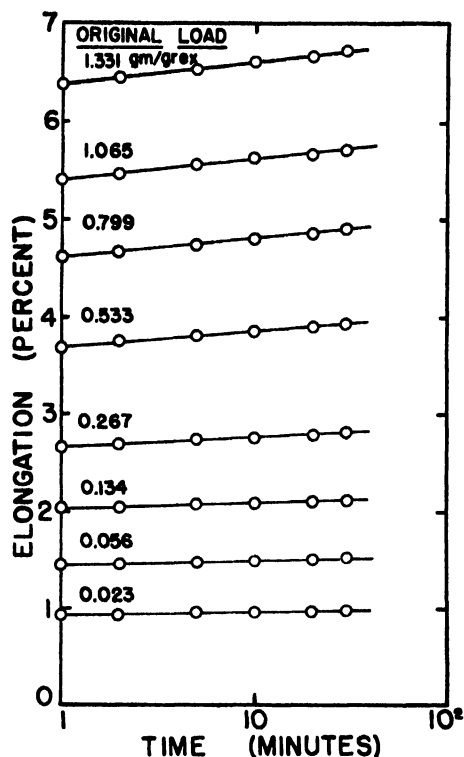


FIG. 2. Creep in a "hot-wet-stretched" 17/4/3 cotton cord.

As a measure of the constant rate of viscous flow in a material or product, the parameter a has technical importance. The comparison of values of a for different materials under equal loads will indicate which will offer the most resistance to continuous yielding or growth in service.

Reference was made above to the anomalous behavior of cellulose acetate strips and Nylon filaments with reference to Eq. (1a). A decelerated creep at the longer times was observed in these materials, contrary to behavior predicted by Eq. (1a). In the light of the foregoing analysis, the observed behavior indicates extensive re-orientation of the chain molecules and enlargement of crystalline areas: an excessive strain-hardening effect. Such an effect might be expected in the linear high polymers, but would not be likely to be found in polycrystalline metals having a granular structure. Presumably, under the action of high stress, or because of easy susceptibility to orientation, the strain-hardening in the acetate and Nylon becomes so pronounced as to alter seriously the viscous

properties of the material and depress the value of the parameter a . In the case of the cellulose acetate the strain over the anomalous region was not much below that at which fracture occurred. The strain-hardening in this region may be regarded as the preliminary stage of the rupture process, and abnormal effects are to be expected. This downward curvature of the *creep vs. log t* curve at high limiting values of t has a counterpart in the complete graph of Eq. (5).²³

CREEP AND RECOVERY IN TIRE CORDS

The measurement in the laboratory of creep and other visco-elastic effects in tire cords provides a useful insight into the "growth" behavior of the cords in a tire in service. Creep experiments with a constant load would appear to parallel closely the conditions in a tire which lead to growth. The cords in an inflated tire on the wheel are continually under tension. The rubber encasing the cords contributes very little toward supporting the load on the tire, and limiting the plastic extension of the stressed cords. In this respect, conditions in a tire differ from those of stress-relaxation experiments of the type in which provision is made to hold the specimen at constant length.

Since the experimental phase of the present study was intended primarily as an exploration for constant-rate flow in tire cord, relaxation experiments have not been included. Relaxation experiments should, however, provide valuable data complementary to those on creep, when the plastic behavior of a particular cord is being exhaustively examined. In tensile creep, the decay of stress is accompanied by an increase in strain tending to offset the decay. By eliminating strain as a variable, the experiments at constant length provide a means for studying the stress history of a cord directly. The observed stress-relaxation in a cord is presumably a reflection of the adjustments taking place at the micellar and molecular levels, without the complication of concomitant deformation in the fiber substance. Furthermore, the glide of fiber on fiber, which is conceivable as a component of the observed creep in cords composed of staple fibers, is substantially eliminated in stress-relaxation experiments.

Creep measurements under constant load were made at room temperature on cotton tire cords

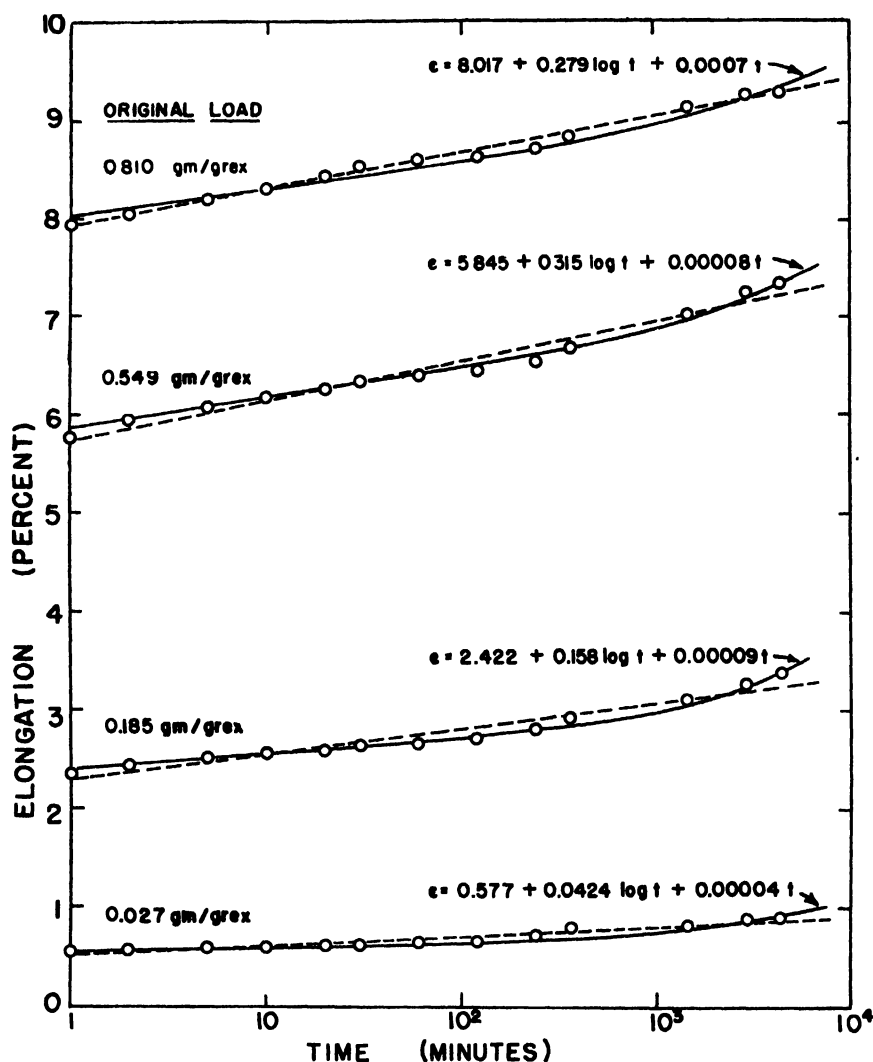


FIG. 3. Creep in a regular, wet-twisted 29/5/3 cotton cord.

of three different types and on a rayon cord. In general, for each cord sample at each of the several loads employed, the time-sequence of observations was made on five specimens. The mean values of the lengths of the specimens at selected times were used in plotting the *elongation vs. time* curves. The observations on the cords, which were hung in a vertical position, were made by means of a cathetometer. The experiments were conducted in an air-conditioned laboratory in which the relative humidity was about 50 percent. The results were not intended for purposes of rigorous comparison between these cords and those tested elsewhere at different times, so extensive equipment for close humidity

control was not set up. In most cases, tests on individual specimens of a particular sample were staggered in such a way that observations corresponding to the same duration of creep were not made simultaneously. Thus, while moderate upward and downward drifts in the relative humidity and temperature may have had an influence on the creep behavior of individual specimens, it is believed that their effects on the measurements cancel out completely enough in the sample averages to justify the main conclusions of the present exploratory study.

The results of these measurements are shown graphically in Figs. 2 to 5. In each figure the original load, based on the cord number (or

specific linear weight) of the sample before loading, is given on the left-hand side of the figure, adjacent to the appropriate curve. In Figs. 2-4 the elongation ϵ is expressed in percentage of the original length of the unstressed cord.

Figure 2 shows the creep in a cotton cord of 17/4/3 construction under a series of loads in which the maximum is about 60 times the minimum. This cord had been given a treatment in which it was stretched in a wet condition, and then dried while still under tension. Since the creep tests extended over a period of only thirty minutes, no constant-rate flow is discernible. The behavior conforms to Eq. (1b). The graph reveals two features having general application to creep in textile materials: *viz.*, the initial deformation, ϵ_1 , is clearly not proportional to the original load, or even to the initial load σ_1 , nor is the value of the coefficient b , in the transient term, proportional to these quantities. In the simplified theories of creep, the quantities corresponding to ϵ_1 and b are often represented as being proportional to σ_1 . Typically, Fig. 2 indicates the invalidity, from the empirical standpoint, of such generalizations. Dillon and Prettyman¹⁵ have noted a similar lack of linearity between the slope b and the load, in their creep data on tire cords.

The results given in Fig. 3, for a regular wet-twisted 29/5/3 cord, cover a much longer period of creep (4320 minutes) than do those given in Fig. 2. The creep behavior here shows a con-

sistent deviation from Eq. (1b). The broken lines are the graphs of equations of the type of Eq. (1b) determined by the method of least squares, while the solid lines are for the best fitting equations of the type of Eq. (1a). The particular numerical equations giving ϵ in percent are shown. The experimental points for the loads of 0.185 and 0.549 g/grex fall in better agreement with the latter curves than with the straight lines. For the remaining two loads one curve shows as good conformity with the observations as the other. All four sets of data, however, show a tendency toward greater elongations at the later time than would be predicted by the best-fitting straight lines drawn through the first eight or nine points in accordance with Eq. (1b). The constant creep rate a tends to increase with the load.

That the constant-rate component of the observed creep in Fig. 3 arises from slippage between fibers is a reasonable tentative inference. If such slippage is the cause, the constant-rate flow should be absent from creep in continuous filament cords, for here the linear coherence does not depend on inter-filament forces. To examine this point, creep observations were made on a commercial viscose rayon cord of 1100/2 continuous-filament construction. The observations, extending over nearly 13,000 minutes are summarized in Fig. 4. It will be readily seen that, though a very moderate load (710 g, or 0.290 g/grex) was used, a constant rate component is in evidence. It becomes appreciable, however,

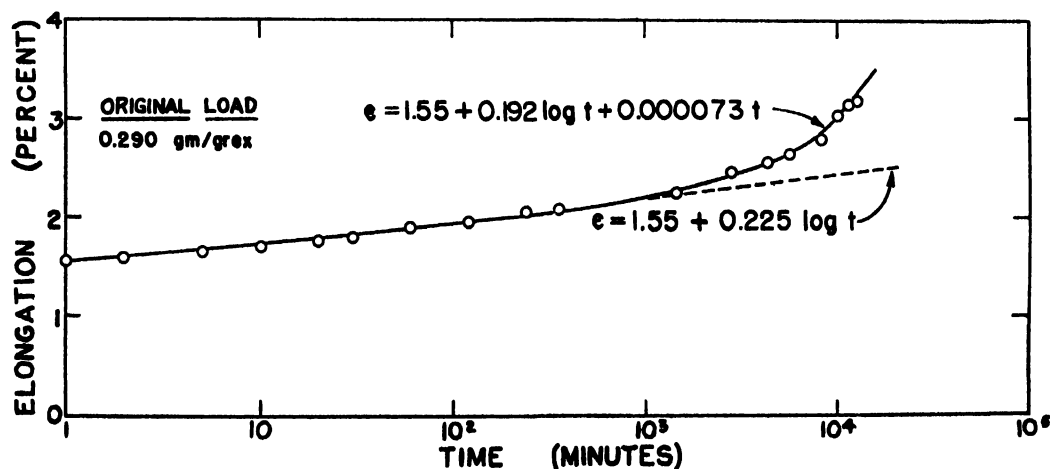


FIG. 4. Creep in a continuous-filament 1100/2 viscose rayon cord.

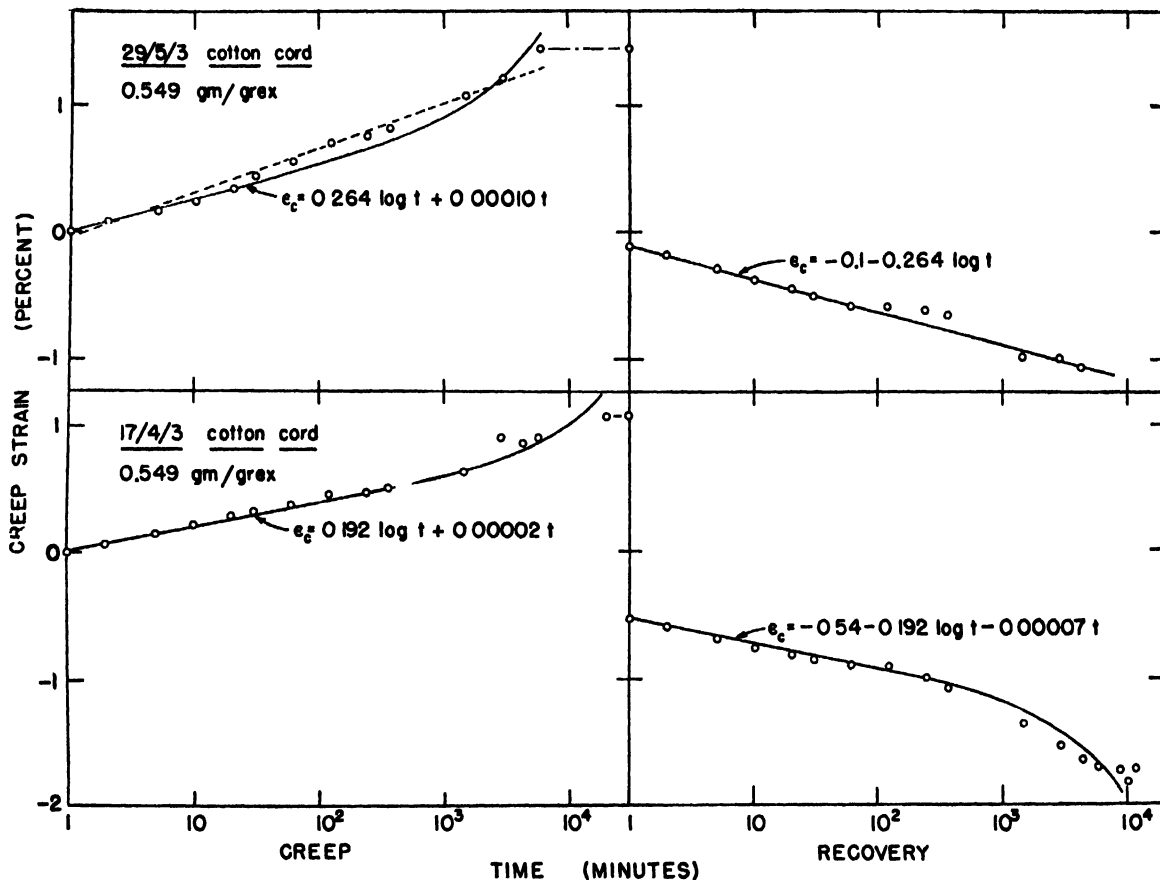


FIG. 5. Creep and recovery in a regular 29/5/3, and a dual stretched 17/4/3 cotton cord. Creep in both cords took place under an original load of 0.549 g/grex.

only after about 1000 minutes of creep. Up to about 1440 minutes (1 day), the elongation is well represented simply by Eq. (1b), as indicated by the dashed straight line. The continuous growth, which in this continuous-filament cord may be confidently taken as arising from viscous flow, proceeds at a slightly higher rate than in the cotton cord in Fig. 3, under a comparable load.

While the extensive experiments of Dillon and Prettyman¹⁸ on cotton and rayon cords fail to show the upward curvature in the graphs of *elongation vs. logarithmic time*, the reason for this would appear to be simply the limitation on the duration of their tests. Most of their tests covered a period of about one day. As has been noted, in the present room temperature tests the constant rate flow did not become noticeable until after about 1000 minutes had elapsed; presumably such a lower time limit would apply

also to elevated-temperature tests such as Dillon and Prettyman's.

In Fig. 5 are shown recovery, as well as creep data, on two cotton cords, one the 29/5/3 regular cord to which Fig. 3 refers, and the other, a 17/4/3 dual-stretched cord, treated by the process of Philipp and Conrad.²⁰ On completion of the creep observations, the loads were removed, except for a clamp which weighed about 30 g, and carried the index mark. The subsequent contractions of the cords were measured periodically to yield the recovery curves in Fig. 5. To condense the graphs, only the creep ϵ_c , i.e., the sum of the terms for the transient and constant rate flows, has been plotted. The points represent the means of the creep in each specimen, expressed as percentage of the particular initial

²⁰ H. J. Philipp and C. M. Conrad, J. App. Phys. 16, 32 (1945).

(1-minute) length of the specimen in the creep test.

The data on the 29/5/3 cord appear to support Eqs. (1a) and (1b) equally well, for both the creep and recovery processes, if the parameters a and b are given negative signs in the application to recovery. Similarly, the creep data on the 17/4/3 cord favor neither equation, but those for the recovery in this cord show a definite conformity to Eq. (1a). The Boltzmann superposition principle¹³ does not hold for these results since very little of the initial strain was recovered when the cords were unloaded. However, it is found, in partial agreement with this principle, that a single value for the coefficient of the transient term $b \log t$ can be used in the equations for both creep and recovery in both cords, to yield curves which satisfy the data moderately well. The necessary values for the rate parameter a are not found to be the same; in one case, that of recovery in the 29/5/3 cord, $a=0$, if b is given the value 0.264, which the creep data favor. A smaller value of b which, with $a=0$, would give a better fit to the recovery data, would lead to a poorer fit in the creep curve. The particular equations and the corresponding curves are shown in Fig. 5.

The appearance of a constant-rate term in the expression for recovery may be deduced from the Eyring theory. In recovery there is no constant external load producing continuous flow. However, the molecular network structure, such as exists in cellulosic fibers, provides small retractive stresses, which come into play when the external load is removed. When the stresses become very small, the rate of deformation, according to the Eyring equations, becomes sensibly constant. The tendency of recovery to approach a constant rate of flow after long intervals of time has been noted also in phenolic plastics by Telfair, Carswell, and Nason.⁴

SUMMARY

As has been shown in the foregoing discussion, there is abundant evidence for accepting Eq. (1a) as the basic, empirical law for creep under constant load. Equation (1b) is merely a special form of Eq. (1a), for the case of $a=0$. Equation (1a) has an adequate foundation in the reaction-

rate theory of deformation, through which also it is related to the empirically-sound Eq. (2a) for stress relaxation. As with other general laws (Newton's viscosity law, for instance), there are materials which behave in an anomalous manner with respect to Eq. (1a), and there are limits outside which the equation does not hold. It evidently does not apply at high stresses near the breaking load, for instance. Notwithstanding their limitations, Eqs. (1a) and (2a), together with their underlying general theory, provide the framework of what may be regarded the *normal* flow behavior of amorphous solids. As the basic, empirical relations for normal elastoviscous phenomena, Eqs. (1a) and (2a) would appear to be in the same category as Hooke's law for pure elastic and Newton's law for pure viscous phenomena.

The cited work of Busse and co-workers, Dillon and Prettyman, and Wakeham, as well as that reported in the present exploratory study, all point to the applicability of Eqs. (1a) and (2a) to the normal plastic behavior of tire cords. The equations have the usual limitations of empirical, semi-theoretical relations, in that they cannot be used validly to compute strains or stresses for periods 100 or 1000 times as long as that over which the constants were evaluated. When Eq. (1a), with the constants for tire cord, is used in such extrapolations, absurdly large elongations are obtained. It must therefore be concluded that the constant-rate flow, represented by the parameter a , does not go on indefinitely, but that after creep times on the order of 10^5 minutes the structure of the cord or its component fibers (as in the case of the acetate and Nylon samples mentioned previously) becomes sufficiently stabilized to counter-balance the residual forces producing the constant-rate flow.

In view of their wide verification, extending to non-textile materials, it would appear feasible to use Eqs. (1a) and (2a) as the basis for more or less standardized tests of creep and relaxation in cords and yarns. Just as a sample of steel is characterized by its moduli of tensile elasticity and rigidity at a certain temperature, so the plastic, or growth, properties of a cord may be identified by the values of the parameters ϵ_1 , a , b , σ_1 , and β under specified conditions of

load, temperature, and humidity. Dillon and Prettyman¹⁵ have taken a step in this direction with their use of the quantities "initial compliance" and "weighted creep." There can be no doubt that the utility, for purposes of comparisons, of creep and relaxation data on tire cord and other textiles, will be broadened if they are interpreted in terms of generalizations which are valid and at the same time are gener-

ally recognized. Equations (1a) and (2a) may be adopted as such generalizations.

ACKNOWLEDGMENT

For such original observational data as are presented in this study, the author has relied on the experimental assistance of Mary L. Nelson, Hilda M. Ziifle, and Trinidad Mares, of whom the first two named also made many of the preliminary calculations.

Fatigue Failure of Rayon Tire Cord

R. . WALLER AND W. E. ROSFVEARE

Viscose Rayon Research, E. I. du Pont de Nemours and Company, Richmond, Virginia

(Received February 16, 1946)

Tire-cord fatigue tests are characterized and interpreted in terms of the following parameters: magnitude and frequency of the cyclic stroke, average load, minimum load, temperature, and humidity. By varying one test parameter at a time, it is shown that the effects of creep and degradation are minimized by using a high frequency and a high stroke. The relative rating of two samples of rayon depends on the parameters of the test and on the modulus and twist structure of the samples. Fatigue breakdown is interpreted in terms of progressive fracture under localized strains occurring at the peak stress of the cycle. A comparison of the fatigue behavior of rubbers, metals, and tire cord suggests they may all involve the same basic process of crack growth.

INTRODUCTION

UNDER repeated cyclic stresses, most solids show mechanical breakdown under stresses far below the normal breaking values. This failure of materials under cyclic deformations is called fatigue or dynamic fatigue, while the

failure of a material under steady load is called creep-failure or static-fatigue.

Since 1900, many papers have appeared on the fatigue of metals. More recently, failures in plastics, elastomers, and fibers have been shown to result from continued cyclic deformations. The fatigue failure of tire cord is responsible for a type of tire failure important in certain classes of service.

The present paper discusses some aspects of the general theory of fatigue and the behavior of rayon tire cord in various laboratory fatigue tests.

FATIGUE TESTS

The rate of fatigue is affected by the structure of the sample, presence of chemically reactive materials, humidity, temperature, period, and magnitude of the localized cyclic strain, and the local magnitude of the minimum strain at the end of a cyclic stroke.

Figure 1 shows the strain throughout the cycle

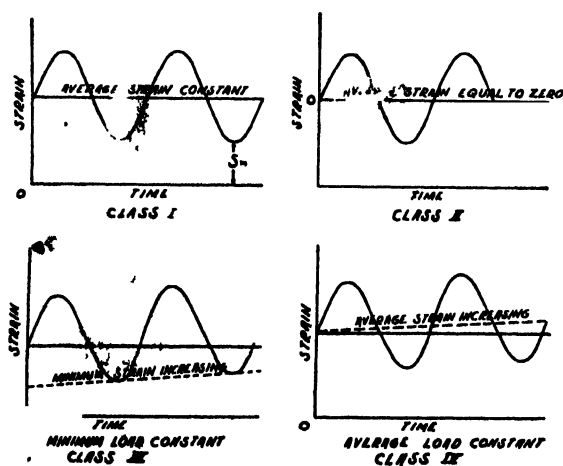


FIG. 1. Strain cycles for various fatigue tests.

for four classes of constant-stroke fatigue tests. These classes differ in the ratio of cyclic strain to average strain and in the changes of average strain from one cycle to the next. These classes of fatigue tests are different from the tests with constant cyclic stress, as usually used for metals. The Class I test of Fig. 1 involves a constant cyclic elongation with a constant minimum strain, S_m , throughout the fatigue life. The Class II test, having equal compression and elongation, is a special case of Class I in which S_m is negative and equal to one-half the cyclic displacement.

Instead of keeping the minimum strain constant as in Classes I and II, one may adopt a Class III test by taking up the slack in a sample under a constant minimum load by means of a ratchet mechanism operating at the point of minimum strain in the cycle. The cyclic deformation remains constant, while the total length of the sample increases, as shown by the increase in minimum strain for Class III of Fig. 1.

Another way of taking up the slack is to hang a weight on the sample, while vibrating the other end up and down with the weight remaining practically stationary. This procedure maintains a constant average load, which likewise causes the sample to grow in length, as shown for Class IV of Fig. 1. The essential difference between the tests of Classes III and IV is in the minimum load being constant in Class III, while the maximum load is constant in Class IV; however, the samples in both tests grow in length until failure.

As either the minimum load in Class III or average load in Class IV is increased, these tests approach a static-fatigue or creep-failure test. As the stroke is increased, these tests become more sensitive to elongation characteristics or elastic limit of the material. A stroke may be chosen that will break a low elongation material in the first stroke and yet not break a high elongation material in many thousand cycles; under such conditions the test will cease to be a fatigue test for the lower elongation material. Therefore, the use to be made of the results of the test must be considered in selecting the conditions for a fatigue test.

A fatigue test may have two purposes, namely, an over-all fatigue rating under service condi-

TABLE I. Fatigue tests.

Test	Class	Stroke % elong.	Load in grams per denier	Temper- ature	% regain moisture in cord	Fre- quency in c.p.m.
A	III	1.6	0.14 (Min.)	49°C	3.7	360
B	IV	1.6	1.0 (Av.)	100°C	2.0	3000
C	IV	0.5	0.65 (Av.)	170°C	Dry	3650

TABLE II. Properties of yarns at 23.9°C (75°F) (3"Z" twist wet or 60 percent relative humidity).

Cord No.	Tenacity g.p.d.		% elongation	
	Cond.	Wet	Cond.	Wet
1	3.62	2.25	10.1	20.3
2	3.70	2.30	9.8	19.6
3	3.56	2.12	9.8	17.9
4	5.23	3.81	7.4	7.8

TABLE III. Properties of cords at 23.9°C (75°F) (1100/2 construction 15"Z"-11"S" twist).

Cord No.	Denier	Conditioned at 60% relative humidity			Oven-dry cord		
		Strength in lb.	% elong. at 10# load	% elong. at break	Strength in lb.	% elong. at 10# load	% elong. at break
1	2492	16.4	9.0	16.5	20.1	4.1	12.1
2	2412	16.0	8.1	13.5	19.0	3.9	10.5
3	2449	15.1	8.1	13.1	17.8	4.0	9.8
4	2439	18.6	5.0	9.4	21.7	2.1	6.4

tions, or a measurement characterizing a sample in terms of what we may call a "true fatigue" resistance which is independent of creep under static load, heat-aging, and deteriorating agents. This true fatigue will be a function of the temperature, moisture content, stroke, and structure of the sample. A complete fatigue rating involves both a rating under standard conditions and the sensitivity of this rating to each of the conditions of which the fatigue test is a function. To obtain practical relative fatigue ratings of materials having widely different properties, such as glass, cotton, rayon, and Nylon, it is necessary to duplicate closely the degree of strain occurring in service failure. Since service conditions are varied and difficult to define, one chooses what appears to be a reasonable set of parameters for a laboratory fatigue test. For narrow classes of materials having similar elongations and temperature sensitivities, etc., such measurements should correlate with service fatigue.

Class I and II tests have been used on rubber¹

¹ S. M. Cadwell, R. A. Merrill, C. M. Slowman, and F. L. Yost, Ind. Eng. Chem. Anal. Ed. 12, 19 (1940).

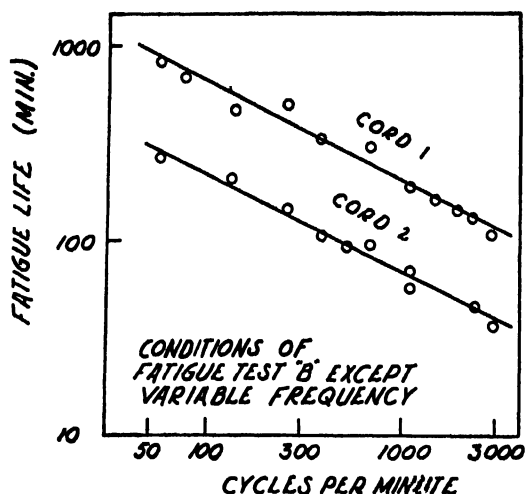


FIG. 2. Effect of frequency of stresses on fatigue life.

and a Class II test has been designed by Ray² for tire cord embedded in rubber. This latter test gives an over-all fatigue rating for the service conditions simulated by the test; however, the results of this test depend on the technique and materials used in embedding the cord in rubber, as well as on the inherent fatigue resistance of the cord materials. Class III and IV fatigue tests have been developed for tire cord by Castricum³ and Busse and others,⁴ respectively. Class I and II tests are much more difficult to apply to tire cord free of rubber than the Class III and IV tests. To simplify both the technique and the interpretation of the results, the present investigation has been confined to the Class III and IV tests.

Experimental Tests

The conditions of some of the selected standard tests are given in Table I. The load is in grams per denier, and the frequency is in cycles per minute.

The effect on fatigue life of varying each of these parameters except stroke, one at a time, has been studied for one or more of these fatigue tests in order to find conditions where the fatigue life is independent of heat-aging and least dependent upon creep under static load. The

² F. Ray, U. S. Patent No. 2,235,622.

³ M. Castricum, U. S. Patent No. 1,923,296.

⁴ W. F. Busse, E. T. Lessig, D. L. Loughborough, and L. Larrick, J. App. Phys. 13, 715 (1942).

properties of the cords used in this study are given in Tables II and III.

All tests were made on a Scott IP4 tester using a loading rate of 16 g.p.d. per minute. Cords 1, 2, and 3 represent three different types of commercial viscose rayon tire yarns, while Cord 4 represents a low elongation viscose rayon yarn.

The coefficient of variation for 54 determinations in Test A was 20 percent. For a normal distribution and a probability of 90 percent, this corresponds to an uncertainty of 18 percent for six observations and 11 percent for 12 observations. Most values reported in this paper represent averages of either six or 12 observed values.

EXPERIMENTAL RESULTS

Effect of Frequency of Stresses

The effect of the frequency of cyclic stresses on the fatigue life (time to failure) of Cords 1 and 2 may be seen in Fig. 2, where frequency and fatigue life are plotted on logarithmic coordinates. A straight line has been drawn through the points from 60 to 3,000 c.p.m. The relationship represented by the straight line may be written in the following form

$$F = k\omega^{-m}, \quad (1)$$

where F is fatigue life in minutes, ω is frequency, and k and m are constants. The fatigue life is proportional to the frequency raised to the negative fractional power $-m$; the number of cycles to failure is equal to the product of the life and the frequency and, therefore, proportional to the frequency raised to the positive fractional power $1-m$. Busse and others⁴ re-

TABLE IV. Irreversibility of fatigue processes in fatigue test A.

Initial fatigue time (hours)	Intermediate treatment	Fatigue time after treatment (hours)	Total fatigue life (hours)
4.7	Control: fatigued direct to failure	3.6	8.3
4.7	Conditioned 24 hours at 60% R.H. 23.9°C	4.2	8.9
4.7	Wet out, dried on jig at original length	3.4	8.1
4.7	Heated for 1.5 hours at 150°C in a dry evac. tube	3.8	8.5

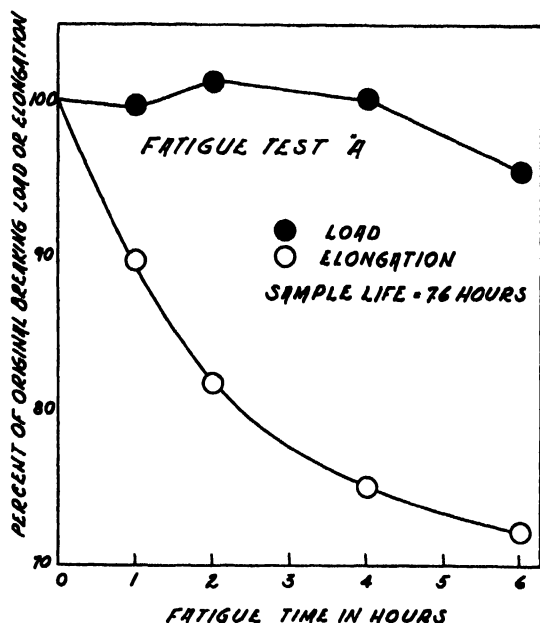


FIG. 3. Effect of fatiguing on breaking load and elongation of a tire cord.

ported the fatigue life of rayon and cotton cords to be independent of the frequency; this was due probably to the heat-aging involved in their test and to the limited frequency range wherein the stress-strain behavior of their tester was unaffected by resonance. No dispersion or critical frequency ranges have been observed, indicating that no critical relaxation frequencies of the cellulose occur in the frequency range employed in this study. The values of m in Eq. (1) are the same for Cords 1 and 2 within the limits of experimental error. Increase in the fatigue life with increasing frequency was reported for rubber by Yost,⁶ although the functional relationship was not given.

Irreversibility of Fatigue Processes

The fatigue processes involved in Test A appear to be irreversible. Samples of Cord 3 having an average fatigue life of 8.3 hours were fatigued for approximately one-half of this time, given an opportunity to recover, and then fatigued to failure. The total time of fatiguing and the kind of intermediate recovery treatment used are given in Table IV. None of the fatigue

resistance expended in the initial period was recovered by the various treatments tried; the differences in total fatigue life are within the experimental error. Busse and others⁴ have shown that the processes of Test C are not reversed by treatments such as mechanical flexing, conditioning, or wetting with various solvents.

The effect of fatiguing, for various times short of failure, on the breaking load and breaking elongation of a tire cord having a fatigue life of 7.6 hours is shown in Fig. 3 for Test A. There is no appreciable change in the breaking load until near the end of the fatigue life; however, the breaking elongation rapidly decreases during the first part of the fatigue life.

Growth during Fatigue Testing

The repeated deformation of a cord results in the production of a permanent elongation. A continuous record of elongation *vs.* time⁷ is given in the upper, left part of Fig. 4 for a sample of commercial tire cord in Test A. The cords grow very rapidly in the first few minutes, after which the rate of elongation becomes proportional to the logarithm of the time as shown by the straight line part of the curve in the lower right part of Fig. 4. The stage just before failure is characterized by a period of rapid growth. Similar straight line plots are obtained for a large part of the creep of a cord under static load. Leaderman⁸ has observed similar behavior

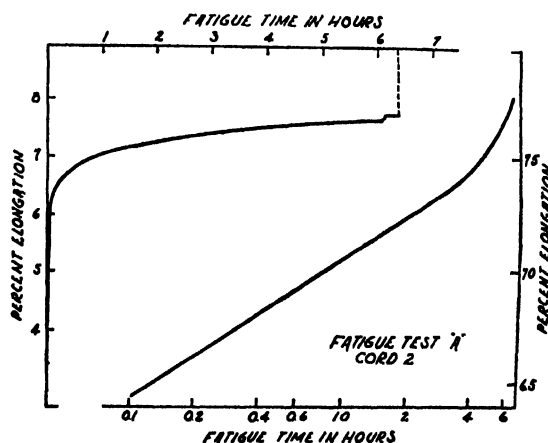


FIG. 4. Growth behavior during fatiguing.

⁷ H. Leaderman, *Elastic and Creep Properties of Filamentous Materials and other High Polymers* (The Textile Foundation, 1943).

⁶ F. L. Yost, *Trans. A.S.M.E.* 65, 881 (1943).

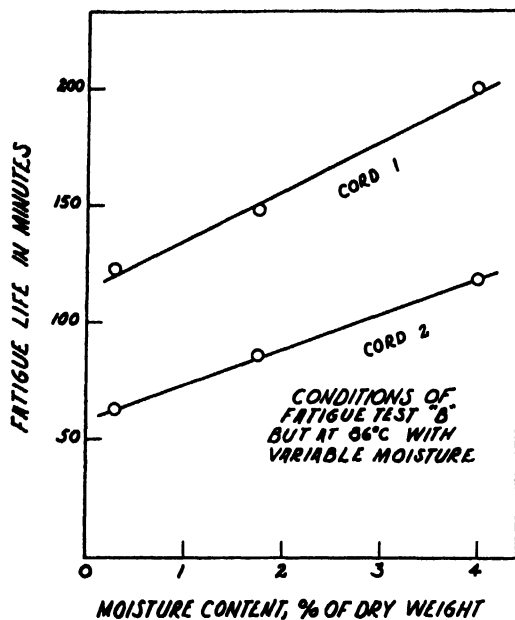


FIG. 5. Effect of moisture on fatigue life.

of the creep under static load for various textile fibers while Chasman⁷ has obtained such creep curves for plastic sheets.

Effect of Moisture on Fatigue

Under conditions approaching those in fatigue Test *B* where heat degradation is insignificant the fatigue life is increased by increasing the moisture content of the cord. Figure 5 shows a plot of the fatigue life of two types of tire cord as a function of the moisture content of the cord obtained on desorption. This effect of moisture is the opposite of the effect of moisture on the life of a cord under a static load. The increased fatigue lives at the higher moisture contents probably result from the lower cyclic stresses present in moist cord. Increasing the moisture content of cord in a tire is undesirable, since it increases the rate of growth of the tire.

• Effect of Temperature on Fatigue

The effect of temperature on life has been determined with conditions other than temperature as in Test *B* and Test *C* and for creep failure under two different static loads. It was impractical to surround the cords with completely dry air or to provide a constant moisture

content in the cord at each temperature. The air supplied to the testers for these temperature coefficient studies contained 60 percent relative humidity at 23.9°C with other conditions as in Test *C* or 13 percent relative humidity at 23.9°C with other conditions as in Test *B*. As the temperature of the test is lowered, the moisture content of the cord will increase. Only below 80°C will sufficient moisture be present to affect the life with other conditions as in Test *B*.

The results of these temperature-coefficient measurements are given in Fig. 6. The lives in both the creep and fatigue tests are exponential functions of the reciprocal of the absolute temperature. For each fatigue test, the temperature coefficient is nearly the same for the two types of high-tenacity rayon yarns.

It follows from the Arrhenius equation giving the activation energy of a physical or chemical molecular rate process⁸ that

$$\log_e L = (E/RT) + C, \quad (2)$$

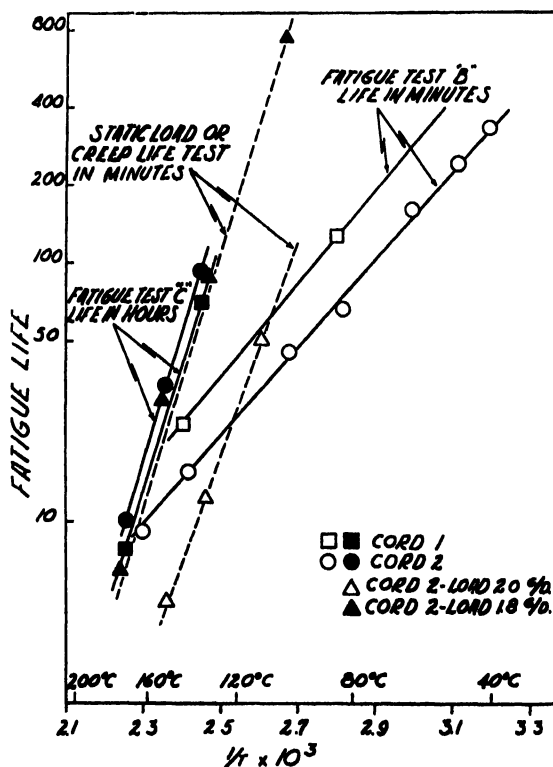


FIG. 6. Effect of temperature on fatigue life.

⁸ S. Glasstone, K. J. Laidler, and H. Eyring, *Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941).

⁷ B. Chasman, *Modern Plastics* 21, 6, 145 (1944).

where L is the time required for a selected amount of change to take place, E is the activation energy for the process, and C is a constant. Table V gives the activation energies determined from the slopes of the straight lines in Fig. 6. The value of the activation energy reported in the table for the fatigue test of Busse and others⁹ is a corrected value calculated from the original data.⁹ This test approaches the conditions of Test C, and the activation energies of these two tests are in reasonable agreement.

Table V also gives a rough activation energy determined in our laboratory¹⁰ for the degradation of Cord 2 in an oven containing air. This activation energy was calculated using the times for equal loss in strength on heating rayon cord at 160°C and 185°C in an oven. The difference between the activation energies for the two creep tests is within experimental error. The activation energy for Test C agrees, within experimental error, with values for creep and with the value for the thermal degradation of tire cord under conditions similar to those in Test C. On the other hand, the comparatively low activation energy found for fatigue Test B indicates that the most important process occurring in this test is not the same as that involved in either heat-aging or creep under static load. In addition, the large difference between the activation energies for Tests B and C suggests that their rates are determined by basically different processes.

The conditions for Test C are more likely to involve heat-aging than the conditions of Test B, since the time available for aging at the high temperature is larger by 600 percent. The importance of heat-aging and creep in Test C cannot be reduced by merely lowering the

temperature, since the temperature coefficient for this test is the same as that for heat-aging and for creep. On the other hand, the effects of heat-aging and the creep under static load for

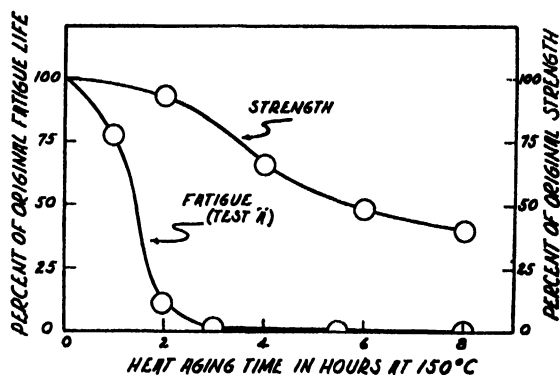


FIG. 7. Effect of heat degradation on fatigue life.

Test B become relatively less important as the temperature is lowered. Since straight lines are obtained for the Test B conditions of Fig. 6, it appears that neither of these factors is important for this test even at the highest temperature studied.

Considerable heat degradation is in evidence in cords fatigued in air under the Test C conditions. This is indicated by discoloration of the test samples. Under the conditions of temperature, moisture, and time in the Type C fatigue test, an unloaded cord may lose 25 percent of its strength because of heat degradation in a normal failure time. Fatigue life is particularly sensitive to previous heat degradation, as shown by the data in Fig. 7. Two hours' heating at 150°C, in sealed tubes containing air and 0.12 gram of moisture per gram of cellulose, reduces the fatigue life by 90 percent, while the loss in tenacity is relatively small. This parallels the behavior of metals where corrosion may greatly reduce fatigue life without greatly changing the tensile strength.

Effect of Load on Fatigue

The life under the conditions of fatigue Tests B and C except for variable load is an exponential function of the average load on the cord. The logarithm of fatigue life is plotted against the average load in Fig. 8. The plot shows that the two samples have the same fatigue life in a Type

TABLE V. Activation energies.

Cord test	Activation energy in calories/kilnetic unit
Fatigue Test B	7800
Fatigue Test C	23000
Fatigue Test of Busse <i>et al.</i>	19000
Creep (1.8 g.p.d.)	22000
Creep (2.0 g.p.d.)	19000
Thermal degradation of tire cord	25000

⁹ W. F. Busse, E. T. Lessig, D. L. Loughborough, and L. Larrick, *J. App. Phys.* 16, 120 (1945).

¹⁰ E. A. Tippetts and A. W. Powell, unpublished data.

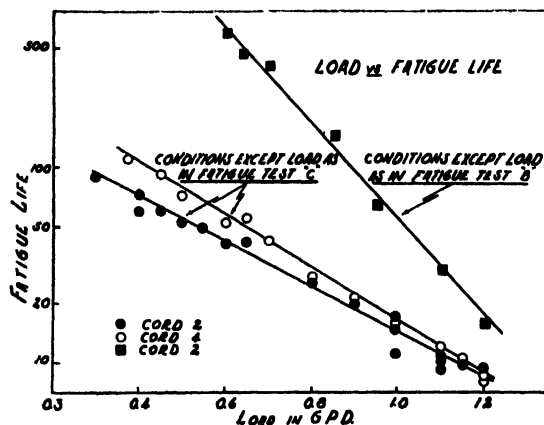


FIG. 8. Effect of load on fatigue life.

C test at an average load of 1.2 g.p.d.; however, the fatigue resistance of Cord 4 is superior to that of Cord 2 at low loads. A change in the relative fatigue ratings of cotton and rayon on changing the load has been reported by Busse and others¹ for a test similar to fatigue Test C. Cords tested at lower average loads are more affected by heat-aging than are those tested at higher loads. This effect may be responsible for the change in relative fatigue ratings observed. Similar transpositions of fatigue ratings at high and low loads are well known in the case of metals. It has been shown for metals¹¹ and cellulose acetate sheets¹² that cyclic stresses below the "endurance limit" of these materials do not produce failure in an infinite number of cycles. At stresses above the endurance limit, the fatigue life is proportional to the applied stress raised to a fractional power. It is possible that if the effects of creep under static load and heat-aging were eliminated, the fatigue life of rayon might follow a similar relationship.

Effect of Stroke and Elongation Characteristics of the Sample

A small change in the stroke makes a great difference in the fatigue life. Changing the stroke of Test A from 1.6 percent to 1.7 percent decreased the fatigue life of a commercial tire cord by 40 percent. Decreasing the modulus of a cord has an effect on the fatigue life similar to de-

¹¹ D. Landau, *Fatigue of Metals* (The Nitralloy Corporation, 230 Park Avenue, New York, New York).

¹² W. N. Findley, *Modern Plastics* 18, 9, 57 (1941).

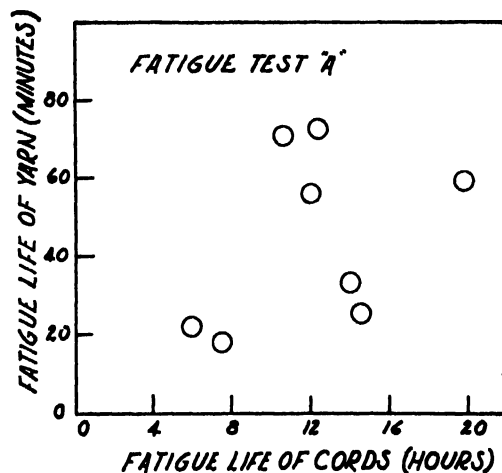


FIG. 9. Fatigue life of yarns vs. cord.

creasing the stroke. The modulus of an 1100/2 rayon cord is usually reported in terms of its reciprocal, the elongation of the cord under 10-pound load. A yarn dried under conditions permitting shrinkage gives a cord with a high elongation under a 10-pound load when compared with a cord obtained from a yarn dried under tension. Such an increase in fatigue life may be considered, in part, artificial, since increasing the 10-pound load elongation above the normal value for a commercial tire-cord yarn is usually accompanied by an objectionable increase in the initial rate of enlargement or growth of the tire under the inflation pressure. For this reason, fatigue ratings of various productions are usually made using cords of equal 10-pound load elongations. If the elongation under 10-pound load deviates a small amount from the specified value, an empirical correction factor may be applied to correct the fatigue to the standard elongation.

Cords with inherently different elongation characteristics will have relative fatigue ratings which depend on the stroke and other conditions of the fatigue tests as Table VI indicates. Cord

TABLE VI. Comparison of ratings on Tests A and C.

Sample	Conditioned elong. at 10# load (see Tables II and III)	Fatigue rating	
		Test A	Test C
Cord 2	8.1	190	90
Cord 3	8.1	100	100
Cord 4	5.0	3	120

4, a high-strength, low elongation viscose-rayon cord, compares unfavorably with commercial tire Cords 2 and 3 in Type A Test, while it is superior in a Type C Test. This appears to be due to the lower stroke of Test C, which measures largely creep and heat resistance.

Fatigue of Filaments, Yarns, and Cords

The fatigue life of a single filament measures its resistance to cyclic stresses along the fiber axis, but when a number of filaments are incorporated into a twisted structure such as a cord, additional possibilities for failure are introduced by the interaction between the filaments. Since tire-cord rayons are highly anisotropic, a fatigue failure resulting from the transverse forces present in a twisted structure may be of a different character than one produced by cyclic stresses along the fiber axis. Single filaments and zero-twist yarn were tested, but the large spread in these results necessitated the use of a low twist yarn (3Z twist) to measure the longitudinal fatigue resistance.

The fatigue life of yarn (3Z twist) is plotted against that of cord (1100/2) in Fig. 9 for the fatigue Test A. This test shows a longer life for the cord than for the yarn, while in the fatigue Test C the yarn has the greater life. The longitudinal fatigue resistance, as measured on a low twist yarn, does not appear to correlate with the fatigue resistance of its cord. This suggests that the relative fatigue ratings of tire yarns may be a function of cord construction. It is well known that the absolute fatigue life of cords from a given type of yarn is also a function of cord construction. The optimum cord construction for one type of rayon yarn is not necessarily the best construction for another type of rayon yarn having different properties.

Effect of Parameters of Test on Fatigue Rating

The ratings of a series of tire cords determined on Test B correlate very well with those from Test A, as shown in Fig. 10. These two tests have strokes with the same percent elongation, but they differ considerably in temperature, frequency, and moisture content of the cord. Figure 10 shows that the results of Test C, using a low stroke and very high temperature, correlate

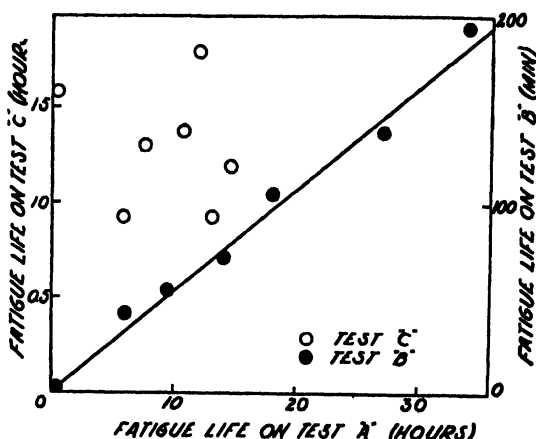


FIG. 10. Relative fatigue ratings with three different tests.

very poorly with the results of the other two fatigue tests. The test parameters to be used in determining a fatigue rating should approach service conditions as closely as possible without making the test life too long. For rayon tire cord, the conditions of the Type B Test give a reasonably short fatigue life in which creep under static load and heat-aging are relatively unimportant. A test combining fatigue, heat-aging, and creep under static load would give more nearly a service rating for a cord if one could make these factors of the same relative importance in the test as in the particular type of tire service for which the tire cord is being evaluated.

DISCUSSION OF THE MECHANISM OF FATIGUE FAILURES

A rayon cord fatigued to incipient failure shows a large number of broken filaments all along the cord. Such a cord shows little loss in strength. Relatively few broken filaments are observed along the length of a cord broken in a tensile tester. These observations suggest that the fatigue occurs in varying degrees at many points in the individual filaments making up the cord. Fatigue failure of the cord results after a critical number of filaments have failed near a point in the cord so that the remaining filaments rupture rapidly because of increased load per filament at this weak spot.

In a cord or yarn it is impossible to watch crack growth in the individual filaments. Some partially broken filaments have been observed in samples of fatigued cords. Fibrillation of the

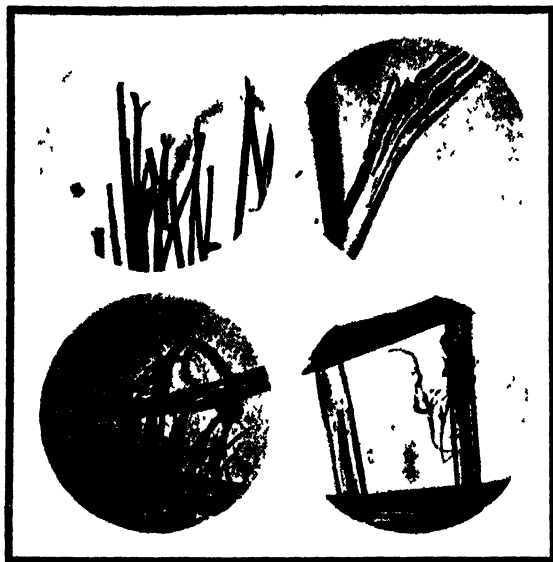


FIG. 11. Photomicrographs of filaments from fatigued cords.

filaments has been observed in fatigued cords in unpublished observations of L. M. Welch in this laboratory. Some of his photomicrographs of filaments from fatigued cords made from typical tire yarns are shown in Fig. 11. Many filaments throughout the fatigued cord show surface fibrillation. Examination of cords removed from the point of maximum fatigue of a tire showing fatigue failure has failed to show this surface fibrillation. This may be due to the difficulty in removing filaments from the partially impregnated cord of a tire and to reduction in the slippage of one filament over another by the adhesive coating which cements the surface filaments together. Cords broken in a tensile tester do not show surface fibrillation and the broken ends of the filaments show less longitudinal fracturing than the ends of filaments from fatigued cords. These observations suggest that fatigue of tire cord is associated with an internal fracturing of the filaments and that surface fibrillation is only important in cords not coated with adhesive.

The initiation of cracks should be more frequent for the higher peak stresses. Increasing the stroke, the average load, or the modulus of the cord reduces the fatigue life. The effect of the latter factor is true whether the modulus is increased by increasing the frequency of the cyclic stroke, decreasing the moisture content of

the cord, or decreasing the 10-pound load elongation of the cord. The use of a fatigue test with a constant peak load, as one uses usually for metals, should reduce the effect of the above variations in modulus on fatigue life.

The temperature coefficient of fatigue is practically zero for most metals. Deviations from this rule have been explained as due to the test sample becoming a different material with different strength properties on changing the temperature.¹³ For normal metals, there appears to be practically no variation of the critical rupture stress with temperature; zinc and bismuth show almost no change in this property from liquid hydrogen temperatures up to room temperatures.¹⁴ A comparison of the behavior of tire cord with metals raises the question as to whether the change in fatigue life with temperature may not be due to a change in the nature of the fiber. The strength of nearly dry rayon cord decreases as the temperature is raised.¹⁵ If fatigue life is as sensitive to strength as it is to average load in Test B, a large part of the change in fatigue life with temperature can be accounted for as being due to change in strength. In addition, changes in the modulus with temperature should affect fatigue life. The available information does not exclude the possibility that the basic process in the fatigue of tire cord may have a zero temperature coefficient as one finds for metals. If the temperature coefficient of the rate determining step is zero for tire cord, then this step must involve the simultaneous rupture of a large number of bonds between aggregates of molecules. These bonds may include one or more of the following types: van der Waals, hydrogen, or chemical bonds. Crack growth by such a mechanism appears to be the most satisfactory explanation covering the fatigue of all types of materials.

The origin of a crack will be the point of maximum microscopic strain. The crack will tend to grow in the direction of the larger adjacent strains. A crack may run into a region of low strain whereupon this crack may cease to grow while new cracks are being formed else-

¹³ M. L. E. Oliphant, *Proc. Roy. Soc.* **171**, 79 (1939).

¹⁴ F. Seitz, *The Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 164.

¹⁵ J. H. Dillon and I. B. Prettymann, *J. App. Phys.* **16**, 159 (1945).

where. A large perfect crystal or even a region of greater order would be a region of low strain, and therefore the presence of such material may retard crack growth. This picture of fatigue predicts the very large observed effect of minute flaws or "stress risers" on fatigue life; it is well known that a high polish on metals greatly increases their fatigue resistance.

In the case of rubber, neoprene, and Butyl B, Cadwell and others² showed that the fatigue life increases to a maximum and then decreases as the minimum elongation of the test is increased. After reaching the maximum, the fatigue life decreases because the effect of increased load is more important than increased crystallization. Fielding¹⁶ has shown that no such increase in fatigue life is found for either Buna-N or GR-S synthetic elastomers. Since this latter group does not show a fiber type of x-ray pattern on stretching, while the first group does, it is probable that crystallization is responsible for the improvement in the fatigue of the first group and that crystals retard crack growth.

Isotropic cellulose and rubber have been shown by Hermans¹⁷ to possess many similarities in their mechanism of deformation. Rayon yarns, on the other hand, are oriented and quite crystalline so that a maximum fatigue life on stretching yarns would not be expected as in the case of rubber. High tenacity viscose rayon may be considered somewhat like stretched rubber, and, like rubber, it will be in the region where increased stretch would rapidly decrease the observed fatigue life by reducing the elastic limit of the substance. Material with a low elastic limit may have excellent fatigue under conditions involving low stroke. A highly oriented rayon may be like rubber at 500 percent

minimum stretch, where rubber has an excellent fatigue life, but on doubling the stroke the fatigue life of rubber is decreased by a factor of over 100. A less crystalline rayon may correspond to rubber stretched 300 percent, where doubling the stroke decreases the fatigue only threefold.

In addition to the gross structural factors affecting fatigue, there are rate factors which affect the behavior of a material in a fatigue test. If a material is very hard, like glass, at normal temperature so that strains are slowly relieved, it will rupture quickly once the elastic limit is exceeded. A very ductile material may be deformed above its elastic limit without fracturing, since the strains are rapidly relieved. Such a material on repeated cyclic stressing under load may be very highly deformed without producing a fatigue rupture, and it is characterized by showing a large creep under high static loads. In some tests it may have a very high observed fatigue life. Rayon is intermediate between very hard and ductile materials.

CONCLUSION

The type of degradation occurring in a fatigue test is determined by the parameters of the test, and varying the values of these parameters may change the order of rating for some samples. Parameters of a fatigue test for tire-cord rayon have been chosen to minimize the effect of heat degradation and creep under a static load.

Metals, plastics, elastomers, and fibers have in their fatigue behavior similarities suggesting progressive fracture as a common basic mechanism for all fatigue failures.

ACKNOWLEDGMENT

The authors wish to acknowledge the valuable assistance of R. E. Spain in obtaining the experimental data for this paper.

¹⁶ J. H. Fielding, *Ind. Eng. Chem.* **35**, 1259 (1943).

¹⁷ P. H. Hermans, *J. Phys. Chem.* **45**, 827 (1941).

Dynamic Characteristics of Rubber Supports from Vibration Table Data

ERNEST G. CHILTON*

Firestone Industrial Products Company, Akron, Ohio

(Received February 11, 1946)

Measurements of the vibration of a free mass which is isolated by a rubber support from a surface vibrating at constant amplitude permit calculations of the dynamic characteristics of the support. This method permits high loadings and the application of considerable force.

INTRODUCTION

THE determination of dynamic characteristics of rubber, both natural and synthetic, has been described by a great number of authors. Values of dynamic rate (or modulus) and internal friction (or damping) in both compression and shear have been determined under many conditions.¹⁻³

However, the usual laboratory apparatus used to find these characteristics in the raw material is rarely capable of withstanding the necessary dead loads, or applying forces of the order of those experienced in service. Gehman⁴ and Kosten⁵ have described methods based on the centrifugal force of revolving eccentric masses.

The method described here utilizes a "vibration table," a horizontal surface made to vibrate vertically with a certain pre-set amplitude and

continually variable frequency. Upon this surface is mounted the specimen under test, and it, in turn, supports the desired load. By recording the amplitude of vibration of this load it is easy to determine the transmissibility of the specimen at the particular frequency. Resonance, however, the condition of maximum transmissibility, is usually chosen as the point at which physical characteristics are determined, for at this point calculations become fairly simple. It will be shown in the derivation how both dynamic rate and internal friction can be determined from the values of resonant frequency and load amplitude together with the known constants of the system.

DERIVATION OF EQUATIONS

Although the motion of the table surface is not perfectly simple harmonic, it is close enough to assume it to be of that form.

Let, therefore, the table motion be described by the equation:

$$x_1 = a \cos \omega t,$$

where x_1 = the distance of the table surface from its neutral position at time " t ," a = the amplitude, or maximum value of " x_1 ," and ω = the frequency of the motion.

The force due to the inertia of the load is given by the product of the mass m of that load and its acceleration d^2x/dt^2 , where x is the displacement of the load from its neutral position at any time (see Fig. 1).

The force due to the dynamic rate of the specimen is proportional to the relative displacement of its ends, and thus to the difference $x - x_1$. While the internal friction force is similarly proportional to the difference in velocities,⁶

$$dx/dt - dx_1/dt.$$

* Viscous friction has been assumed in these calculations for reasons of simple mathematical treatment as in references 1-4.

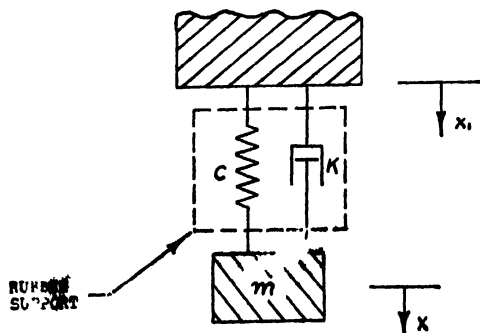


FIG. 1. Schematic diagram of vibration system.

* Now on leave of absence at Stanford University, California.

¹ S. D. Gehman, D. E. Woodford, and R. B. Stambaugh, *Ind. Eng. Chem.* **33**, 1032 (1941); *Rubber Chem. Tech.* **14**, 843 (1942).

² R. B. Stambaugh, *Ind. Eng. Chem.* **34**, 1358 (1942); *Rubber Chem. Tech.* **16**, 400 (1943).

³ J. H. Dillon, I. B. Prettyman, and G. L. Hall, *J. App. Phys.* **15**, 309 (1944).

⁴ S. D. Gehman, *J. App. Phys.* **13**, 402 (1942); *Rubber Chem. Tech.* **15**, 869 (1942).

⁵ C. W. Kosten, *Proc. Rubber Tech. Conf.*, London 987 (1938); *Rubber Chem. Tech.* **13**, 381 (1939).

Hence, with due regard for direction, the equation of forces of the system can be written as:

$$m \frac{d^2x}{dt^2} + k \left(\frac{dx}{dt} - \frac{dx_1}{dt} \right) + c(x - x_1) = 0, \quad (1)$$

where k is the damping or friction constant and c is the dynamic rate or modulus. Substituting for

$$x_1 = a \cos \omega t$$

and

$$dx_1/dt = -a\omega \sin \omega t,$$

we get:

$$m d^2x/dt^2 + k dx/dt + cx = ca \cos \omega t - k a \omega \sin \omega t. \quad (2)$$

The transient part of the solution of this differential equation is of little interest as it describes the first few moments of motion before the steady state has been reached. The steady-state solution, however, is of the form:

$$x = M \sin \omega t + N \cos \omega t. \quad (3)$$

Differentiating this:

$$dx/dt = M\omega \cos \omega t - N\omega \sin \omega t,$$

$$d^2x/dt^2 = -M\omega^2 \sin \omega t - N\omega^2 \cos \omega t = -\omega^2 x.$$

Substituting in (2) and letting ωt assume values of 0 and $\pi/2$, respectively, we get:

$$(c - m\omega^2)M - k\omega N = -k a \omega,$$

$$(c - m\omega^2)N + k\omega M = ca.$$

Using $n = k/2m$ and $p^2 = c/m$ this becomes:⁷



FIG. 2. Typical vibration support with 100-lb. load on vibration table.

⁷ S. P. Timoshenko, *Vibration Problems in Engineering* (D. Van Nostrand Company, Inc., New York, 1928), pp. 23 and 26.

$$(p^2 - \omega^2)M - 2n\omega N = -2an\omega,$$

$$(p^2 - \omega^2)N + 2n\omega M = p^2 a.$$

Solving these equations for M and N :

$$M = \frac{2na\omega^3}{(p^2 - \omega^2)^2 + 4n^2\omega^2}, \quad (4)$$

$$N = \frac{a(4n^2\omega^2 + p^4 - p^2\omega^2)}{(p^2 - \omega^2)^2 + 4n^2\omega^2}. \quad (5)$$

By setting $dx/dt = 0$ it is easily shown that the maximum value of x :

$$x_{\max} = (M^2 + N^2)^{1/2} = \text{amplitude} = A.$$

Substituting for M and N from Eqs. (4) and (5):

$$A = \frac{a(4n^2\omega^6 + 16n^4\omega^4 + p^4 - 2p^6\omega^2 + p^4\omega^4 + 8n^2\omega^2p^4 - 8n^2\omega^4p^2)^{1/2}}{(p^2 - \omega^2)^2 + 4n^2\omega^2}. \quad (6)$$

At resonance, when A is a maximum,

$$p = \omega^8 \quad (7)$$

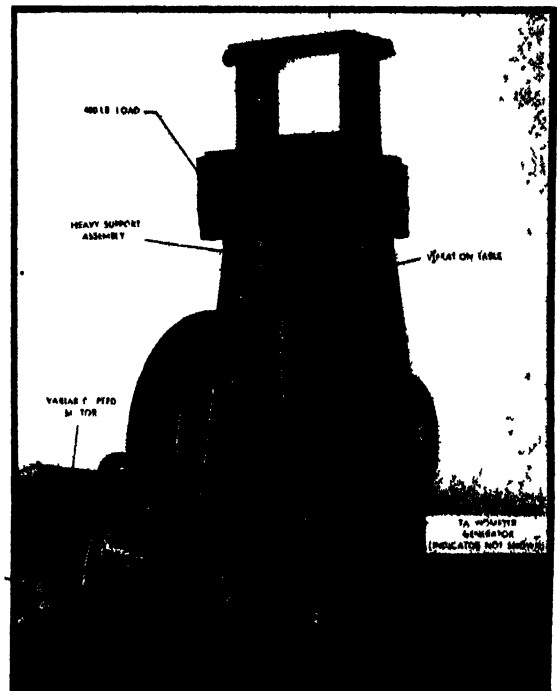


FIG. 3. Vibration support with 400-lb. load in heavy-duty vibration machine.

⁸ This is strictly true only for $n \rightarrow 0$; it can, however, be shown by setting $dA/d\omega = 0$ that $p = \omega$ is in error only by a term proportional to n^4 . Since in most practical cases n is quite small, the assumption is well founded.

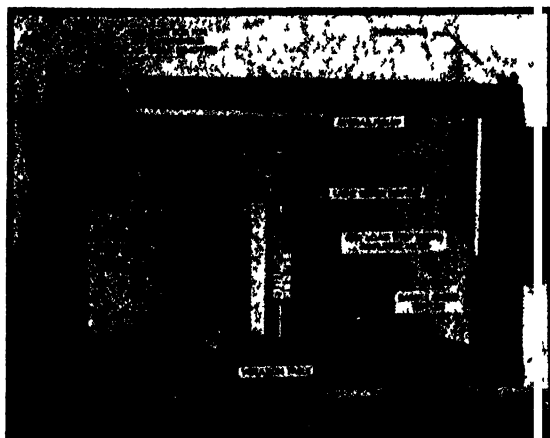


FIG. 4 Small shear type rubber support with 2-lb. load

or, in other words, the natural frequency

$$\omega_n = p = (c/m)^{1/2}$$

or

$$c = m\omega_n^2 \quad (5)$$

Substituting (7) into Eq. (6) we find

$$A = \frac{a}{2n} (\omega_n^2 + 4n^2)^{1/2}$$

Re-substituting:

$$n = \frac{k}{2m},$$

$$A = \frac{a}{k} (\omega_n^2 m^2 + k^2)^{1/2},$$

or

$$k = \frac{\omega_n m a}{(A^2 - a^2)^{1/2}} \quad (9)$$

Thus, with the observed ω_n data, both c and k can easily be determined.

APPLICATION

Considerable use has been made of the above theory in determining the vibration characteristics of automotive and industrial supports. Figures 2 and 3 show typical examples; one, a light engine support tested under a load of 100 lb., the other (Fig. 3) a heavy motor mounting assembly tested under 400-lb. load. Even heavier loads than this are anticipated, but apparatus capable of handling them has not been completed yet.

Figure 4 shows a very light shear type support tested under a load of approximately 2 lb. Typical results obtained with this type of support are shown below. A visual amplitude meter is used to locate the point of resonance and is capable of measuring amplitudes to within 0.001 inch, while a crystal pick-up will give greater accuracies in determining both frequency and amplitude. The output of this pick-up is recorded with an oscillograph.

Figure 5 shows a full view of the light vibration table unit fitted with air conditioning equipment that allows tests to be made at a wide range of temperatures.

TYPICAL RESULTS

The results shown in Table I were gained from a set-up similar to that shown in Fig. 4. Tests were made at room temperature of 25°C.

DISCUSSION

The method described here lends itself well to a speedy determination of dynamic characteristics of rubber supports. The variables of the physical

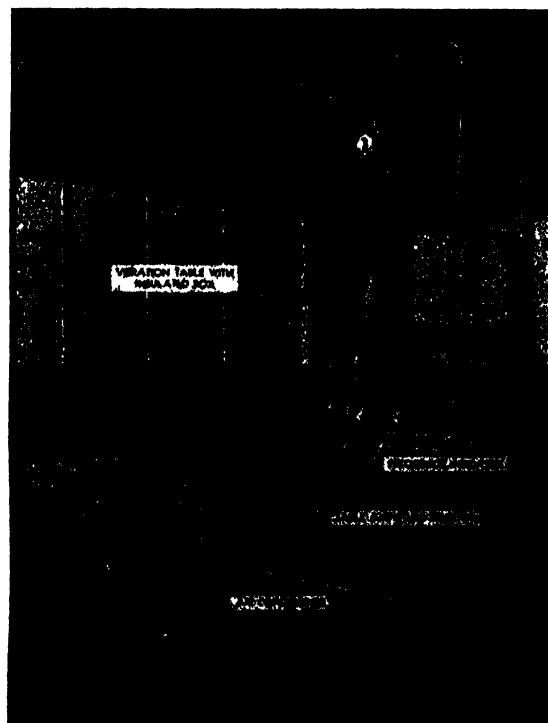


FIG. 5. View of vibration testing equipment.

set-up can be changed quickly and separately to give any desired resonance conditions. They are easily recorded, for displacements, frequencies, and dead weights are the only quantities to be measured. Calculations, also, are simple and straightforward.

A limitation is the fact that it is difficult with the present machines to make tests at constant strain which may, however, be remedied by redesigning the vibration table so that the forcing amplitude can be varied while the machine is running at the resonant frequency of the test piece.

ACKNOWLEDGMENTS

The author wishes to thank the Engineering Research Department of the Firestone Tire and Rubber Company for the equipment used and for permission to publish the photographs. Appreci-

Type stock	Hevea	Buna-S	Neoprene	Butyl	Buna-N
Shore A durometer	40	41	42	38	40
Load (g)	1027	1027	1027	1027	1027
Nat. freq. (c.p.s.)	23.15	27.85	24.3	26.5	26.2
Dyn. rate (p.s.i.)	70	102	77	92	89
Table ampl. (inches)	.0125	.025	.0125	.025	.025
Dyn. strain	.21	.13	.17	.08	.14
Int. friction (kilopoises)	1.99	7.83	2.63	13.18	6.84

ation is also due Dr. J. H. Dillon, and Messrs. E. F. Riesing, G. L. Hall, and A. B. Lane for their advice and cooperation, Mr. P. P. Crisp for his continued interest, and the Firestone Tire and Rubber Company for permission to publish this article.

The Absorption of Microwaves by Gases

W. D. HERSHBERGER

RCA Laboratories, Princeton, New Jersey

(Received December 15, 1945)

The absorption by ammonia of electromagnetic waves having a length in the one-centimeter range has been known for some years. An investigation has been made to determine whether other gases show similar absorption for microwaves and as a result fourteen additional gases have been found whose absorption is comparable to that of ammonia. Among these gases are dimethyl ether, a variety of amines and alkyl halides, and several others. Measurements on the absorption coefficient and dielectric constant of these gases taken at 1.25 cm at room temperature and a pressure of one atmosphere are given. The frequency at which the absorption coefficient attains its maximum value may be inferred from the curve: absorption coefficient *vs.* pressure. Data on the absorption of several gas mixtures are given. Possible molecular mechanisms adequate to account for the large absorptions observed are discussed together with the conclusions reached.

IT has been known for ten years that ammonia strongly absorbs 1.25-cm microwaves. This paper presents the results of a somewhat detailed study of this effect in ammonia and also gives the results of an investigation which was made to determine whether there are gases other than ammonia which show similar absorption.

Among the absorbing gases found are a variety of alkyl halides, dimethyl ether, and various amines. However, the mechanism of absorption for some of these other gases obviously is different from that involved in ammonia absorption.

The order of magnitude of these absorption effects will be made clear by reviewing the pertinent information for ammonia as originally given by Cleeton and Williams.¹ For this gas, at room temperature and at a pressure of one atmosphere, the absorption is such that a plane wave having a wave-length of 1.25 centimeters will be attenuated to $1/e$ of its initial power on traversing a layer of ammonia 1.2 meters thick. Further, the absorption band is relatively wide,

¹ C. E. Cleeton and N. H. Williams, *Phys. Rev.* **45**, 234 (1934).

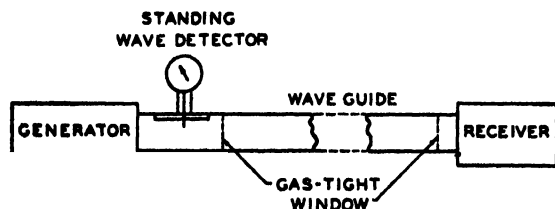


FIG. 1. Wave guide system for measurements on gases.

since under the above conditions absorption falls to one-half of its maximum value at approximately 1.0 centimeter and 1.5 centimeters.

TECHNIQUES

Cleaton and Williams measured absorption in ammonia by passing a parallel beam of microwaves through a cell with a 36-inch by 45-inch cross section such that a transmission path of 16 inches in the gas was traversed. The transmission loss through this cell was observed experimentally on substituting ammonia for air in the cell. Since the time of their investigations, wave-guide measuring techniques² have been developed and were used in these experiments. The power absorption coefficient for unconfined plane waves having a length λ in a given gas is given by

$$\alpha_p = \alpha_g [1 - (\lambda/2b)^2]^{1/2}, \quad (1)$$

where λ_g is the experimentally measured absorption coefficient of the gas which is enclosed in a rectangular wave guide used in the H_{01} mode, b is the width of the guide.

The derivation of Eq. (1) is given in the appendix as well as that of other quantities that depend on the solution of Maxwell's equations in a wave guide containing a dielectric with small losses.

α_g is determined experimentally by noting an initial power reading P_0 in the receiver and a second power reading P , after introducing the gas under study into a length of guide L situated between the source of microwave power and the receiver.

Then

$$\log P/P_0 = -\alpha_g L,$$

and finally

$$\alpha_p = -([1 - (\lambda/2b)^2]^{1/2} \log P/P_0)/L. \quad (2)$$

² G. C. Southworth, *Bell Sys. Tech. J.* **15**, 284 (1936); W. L. Barrow, *Proc. I.R.E.* **26**, 1520 (1938); J. C. Slater, *Microwave Transmission* (McGraw-Hill Book Company, Inc., New York, 1942), Chapter III.

Figure 1 shows in schematic fashion the microwave system used in taking measurements. Components used include a generator, a receiver, a standing wave detector, and a section of wave guide. Various guide lengths in the range from three feet to 30 feet were used. Two mica windows sealed into the wave-guide system were used to confine the gas in a given section of the guide. The reflection introduced by a window was compensated by a reflection set up from a tuning element so the over-all effect was a matched guide at the operating frequency. In measurements in which dielectric constant alone was desired, only one window and tuner of this sort were used while the remote end of the guide was closed by a metal diaphragm, thus giving a pronounced standing wave in the system. The position of a minimum in the standing wave pattern was located just outside the window by the use of a standing wave detector. This minimum was located with the test section of guide evacuated, and the shift in its position observed as the gas under study was admitted to the guide. This technique closely resembles that used in optical refractometers in which a specific interference fringe is kept on a cross hair by means of an optical adjustment. The theory of the method as it applies to wave guides is given in the appendix. If ϵ is the specific inductive capacity of the gas and δ is defined by

$$\epsilon = 1 + \delta,$$

then

$$\delta = \frac{2\Delta L}{L} \left[1 - \left(\frac{\lambda_{00}}{2b} \right)^2 \right], \quad (3)$$

where ΔL is the shift in the position of the minimum in the standing wave pattern, L is the length of the test section of guide, and λ_{00} is the free space wave-length.

A method which permits simultaneous determination of specific inductive capacity and absorption coefficient involves the use of a voltage standing wave ratio of the order of 1.1. Under this condition, Fig. 2 shows the wattmeter reading obtained as a function of the pressure of ethyl chloride in a guide 10 feet long. The maxima in this figure occur when the tuning conditions for maximum power transfer have been satisfied. The cyclic variations in wattmeter reading permit calculation of specific inductive

capacity, while absorption is determined from the behavior of the axis of the curve which is shown as a broken line. This method is best suited for work with gases having relatively high specific inductive capacity but only moderate losses. In any event, it is of the utmost importance to have a closely controlled standing wave ratio in measurements by any of the methods outlined.

RESULTS AND DISCUSSION

It was anticipated that the most promising gases for microwave absorption would be those organic gases having a dipole moment and in which there exists the possibility for hindered rotation of a methyl group with respect to some other group about a single carbon-to-carbon bond. The two equilibrium positions in the ammonia molecule for the nitrogen atom are separated by a potential barrier of about 6 kcal. per mole, while the height of the barrier which separates two adjacent minima in the internal rotation energy curve in ethane is estimated to be about 3 kcal. per mole. The energy level splitting responsible for ammonia absorption at 1.25 centimeters decreases exponentially as the square root of the area of the potential hill cut off by an energy level, so it seemed reasonable to expect that the level splitting in ethane-like molecules, while considerably more complicated than in ammonia, would lead to absorption in the microwave frequency region. Accordingly, the first new gas used in taking measurements was ethyl chloride. This gas was found to exhibit about one-fifth the absorption of ammonia.

TABLE I.

Material	$\alpha \times 10^4$	$\delta \times 10^3$
Ammonia	78.	5.5
Methyl fluoride	10.0	9.2
Methyl chloride	8.25	10.5
Methyl bromide	6.6	9.
Ethyl chloride	14.5	12.
Freon 22 CHClF_2	10.5	5.4
Freon 21 CHCl_2F	10.6	5.8
Freon 12 CCl_2F_2	3.	4.
Dimethyl ether	4.1	4.5
Ethylene oxide	7.	11.
Sulphur dioxide	8.	3.5
Methylamine	9.	4.2
Dimethylamine	7.	3.1
Ethylamine	11.	4.1
Hydrogen sulphide	0.5	2.7

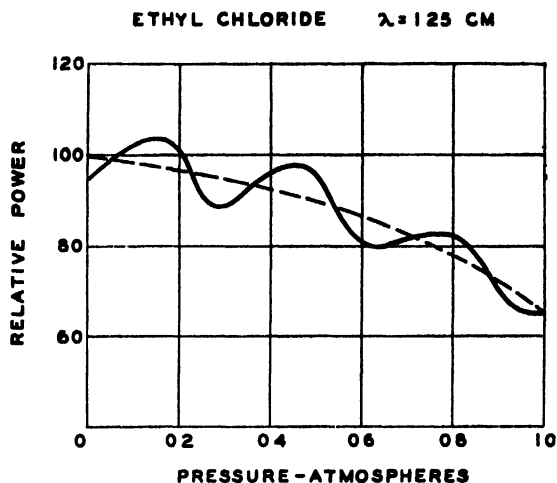


FIG. 2. Receiver power using mismatched wave guide.

Samples of various amines and dimethyl ether were then secured and were also found to absorb microwaves. However, the simple hypothesis involving only hindered rotation was not found to be adequate when measurements showed that the methyl halides and the various gases known commercially as Freon also strongly absorb microwaves. Table I gives the values for absorption and δ for the gases found to be highly absorbing. These values were obtained at room temperature and a pressure of one atmosphere absolute at a wave-length of 1.25 centimeters. The reciprocal of α is the path length in meters which suffices to reduce the power in a plane 1.25-centimeter wave to $1/e$ of its initial value. Thus for ammonia this path length is 1.28 meters while for methyl fluoride it is ten meters. The present measurements indicate that the values reported by Cleeton and Williams for ammonia are perhaps seven percent too high. Measurements on the first ten materials were taken with the 30-foot guide and are accordingly more accurate than the measurements on the remaining five substances where the 5-foot section only was used.

Polyatomic gases whose absorption is either zero or too small for measurement with the present technique include: hydrogen, oxygen, nitrogen, carbon dioxide, carbon monoxide, ethylene, acetylene, nitrous oxide, methane, ethane, propane, propylene, *n*-butane, isobutane, butene-1, butene-2, isobutylene, butadiene, and trimethylamine.

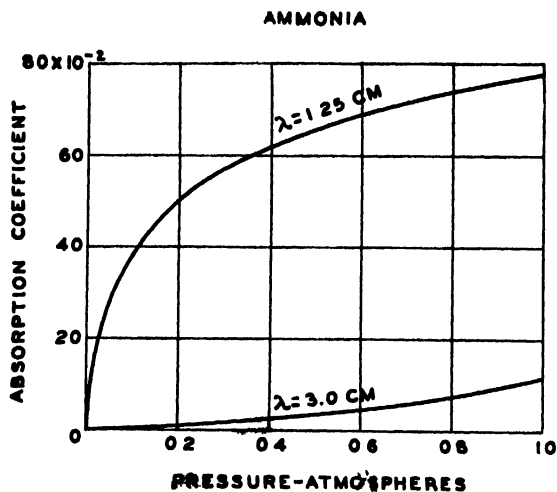


FIG. 3. Absorption of ammonia at 1.25 and 3.0 centimeters.

The smallest absorption which it is possible to detect using a 30-foot section of wave guide is in the neighborhood of 0.5×10^{-2} per meter, which means that measured power at the receiver falls to 94 percent of its initial value on introducing a gas characterized by this absorption coefficient. This is the same change that one would observe in a mismatched wave guide system with a voltage standing wave ratio of 1.03 and which arises because the specific inductive capacity of the gas differs from unity. Thus the problem reduces to one of distinguishing between monotonic received power changes which arise from gas losses, and the cyclic power changes which arise purely from wave guide mismatch.

For low loss materials, δ is less than 10^{-3} and may be as low as 10^{-4} . It will be noted that in Eq. (3) the factor $2[1 - (\lambda_{vs}/2b)^2]$ is of the order of unity, so δ approximately equals $\Delta L/L$. Thus if L is three meters, and δ of the order of 10^{-3} , we must measure ΔL to 3 millimeters, which is feasible. However, if we attempt to increase precision by increasing L to, say, six meters, the standing wave pattern becomes poorly defined owing to copper losses in the guide giving rise to considerable uncertainty in setting the standing wave detector, hence giving us no marked increase in precision.

The location of the maximum in the curve relating absorption and frequency is of great theoretical interest in studies involving molecular structure. The maximum in this curve was deter-

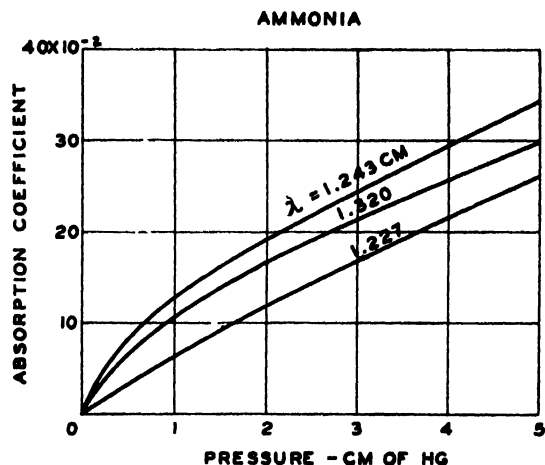


FIG. 4 Absorption of ammonia at 1.227, 1.243, and 1.320 centimeters.

mined for ammonia by Cleeton and Williams by the use of a series of microwave generators and they made a point-by-point experimental determination. A useful method entailing less labor is to employ experimentally determined curves relating absorption and pressure taken at several discrete frequencies and then use the theory of Lorentz broadening of spectral lines to determine the parameters which enable this curve to be constructed. The character of these pressure-absorption coefficient curves for ammonia is shown in Figs. 3 and 4. When the operating frequency is nearly the same as the proper molecular frequency, the curve displays a knee at a relatively low pressure. The effect may be described qualitatively as follows: At first absorption increases very nearly linearly with pressure until a pressure is reached such that collision frequency in the gas becomes comparable with the operating frequency. At still higher pressures, the increase in absorption due to increase in density is very largely offset by the fact that intermolecular collisions limit the length of the absorbed wave trains. The spectral distribution in these interrupted wave trains is such that very little absorption is gained at the proper molecular frequency but absorption at neighboring frequencies only is enhanced. At a frequency well removed from the proper frequency, absorption may increase with the square of the pressure. The curve for 3.0 centimeters in Fig. 3 shows an increase in absorption of this character. The curves in Fig. 4 show how sen-

sitive absorption is to frequency at reduced pressures.

The effect on absorption of impurities was studied by mixing a non-absorbing gas with an absorbing gas, and some results are shown in Fig. 5. Mixtures of ammonia with hydrogen, with oxygen, and with methane were used and no difference in the net effect could be measured which might arise because of the quite wide range in molecular weights of the non-absorbing gases—from two to 32—in comparison to the molecular weight of the ammonia which is 17.03. The solid lines represent the experimental results obtained. The dashed lines which were drawn in later represent the absorption of a constant quantity of ammonia in a given length of wave guide but with an increasing partial pressure of the non-absorbing gas. The curves indicate that the non-absorbing gas has a relatively small yet measurable effect on the absorption of the ammonia; that is, absorption is to a first approximation dependent on the partial pressure of the ammonia. The presence of the foreign gas presumably widens the absorption band at the expense of absorption at its maximum. In these experiments it is essential to use premixed gases; that is, quite misleading results will be obtained if ammonia is admitted to an evacuated guide to some chosen pressure, and the foreign gas is then admitted at one end of a guide section compressing the ammonia before it into one end of the section. Under these conditions the manner in which net absorption changes with time might well be made the basis of measurements in diffusion rates in slender tubes.

Absorption by the methyl halides strongly suggests that we are dealing with transitions between the lowest rotational states of these molecules. However, the calculated absorption due to rotation for these gases is much less than that observed. The theoretical absorption for methyl chloride is approximately the same as that calculated for carbonyl sulphide which is triatomic and linear. The use of this material in a critical check on the theory was suggested by J. Turkevich. However, measurements show that the absorption of carbonyl sulphide is certainly less than one-tenth that of methyl chloride. The small calculated absorption due to rotation arises primarily from the sparsity of the

molecular populations in the lower rotational states at room temperature. Thus but one molecule in 1000 of carbonyl sulphide is in the appropriate rotational state for 1.25-cm absorption at room temperature. Hence, we are forced to look to "inversion doubling" of energy levels in the methyl halides akin to that found in ammonia to account for the experimental results. Further studies designed to evaluate the contributions from all causes to the measured absorption are now under way as well as analyses of pressure *versus* absorption curves for the materials listed in Table I.

APPENDIX: THE EFFECT OF THE DIELECTRIC ON PROPAGATION IN WAVE GUIDES

The H_{01} or TE_{10} solution of Maxwell's equations which satisfies the boundary conditions imposed by the conducting walls of a rectangular wave guide of width b has the form

$$\begin{aligned} E_z &= A \sin(\pi y/b) \exp(j\omega t - \nu z), \\ H_y &= B \sin(\pi y/b) \exp(j\omega t - \nu z), \\ H_x &= H_0 \cos(\pi y/b) \exp(j\omega t - \nu z). \end{aligned}$$

It is, of course, trivial to express A and B in terms of H_0 by the use of Maxwell's equations. However, here the interest centers in the propagation function ν which is complex when we have losses in the dielectric. Each of the components of $\mathbf{H}$ or $\mathbf{E}$ satisfies the wave equation which now includes a dissipation term. We require ν in terms of wave guide dimensions, operating fre-

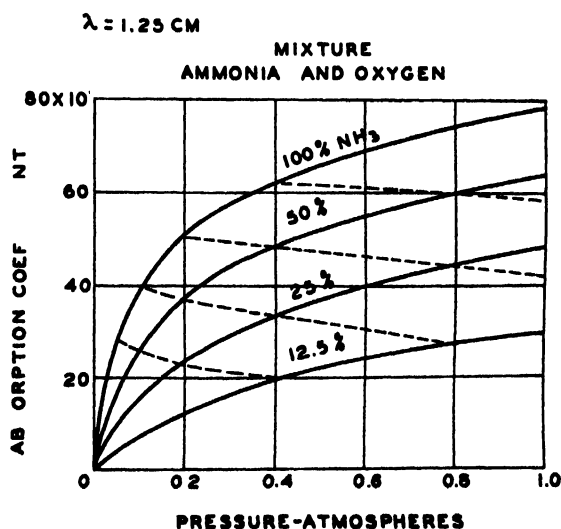


FIG. 5. Absorption of mixtures of ammonia and oxygen.

quency $f = \omega/2\pi$, and the properties of the dielectric. Thus H_z must satisfy the wave equation

$$\nabla^2 H_z - \sigma \mu \dot{H}_z - \mu \epsilon \ddot{H}_z = 0.$$

This is true if

$$\nu^2 = (\pi/b)^2 - \omega^2 \mu \epsilon + j \omega \mu \sigma.$$

If we let $\nu = a + j\beta$,

$$\beta^2 = \omega^2 \mu \epsilon - (\pi/b)^2 + a^2,$$

while

$$2a = \omega \mu \sigma / \beta.$$

β determines wave-length in the guide and the speed with which power is transmitted, which a is the amplitude damping factor. Hence $2a$ is the power absorption coefficient. ϵ is the permittivity of the dielectric inside the guide, μ is the permeability of this dielectric while σ is its conductivity.

In this paper the m.k.s. system of units is employed. Thus for free space, we use the subscript zero and have $\epsilon_0 = 8.85$ micromicrofarads per meter, $\mu_0 = 1.257$ microhenries per meter, while

$$c = [\mu_0 \epsilon_0]^{-1/2}.$$

For media such as those considered here

$$\epsilon = \epsilon \epsilon_0, \quad \mu = \mu \mu_0,$$

where ϵ is the specific inductive capacity and μ is the specific permeability. For attenuation of the order of magnitude encountered in gases, a has a negligible effect on wave-length so

$$\beta = (\omega/v)[1 - (\lambda/2b)^2]^{1/2},$$

where $v = [\mu \epsilon]^{-1/2}$ which is the speed of propagation in the gas for unconfined waves. Also

$$\lambda = v/f,$$

while $\lambda_0 = 2b$ is the cut-off wave-length for the guide; thus there is no wave propagation when $\lambda > \lambda_0$. Then

$$\alpha_g = 2a = \sigma(\mu/\epsilon)^{1/2}[1 - (\lambda/2b)^2]^{-1/2}.$$

But $\sigma(\mu/\epsilon)^{1/2}$ is the power absorption coefficient α_P for plane waves. Or

$$\alpha_P = \alpha_g[1 - (\lambda/2b)^2]^{1/2}.$$

This is Eq. (1).

To derive Eq. (3) dealing with the shift in the

position of the minimum in the standing wave pattern resulting from the admission of gas into a wave guide, we take $\psi = 1$ and again neglect the effect of attenuation on wave-length in the guide.

Then

$$\beta^2 = \epsilon(\omega/c)^2 - (\pi/b)^2.$$

Wave-length in the guide is given by

$$\lambda_g = 2\pi/\beta.$$

Let λ_{rs} represent the free space wave-length in vacuum. Then

$$\lambda_g = \lambda_{rs}[\epsilon - (\lambda_{rs}/2b)^2]^{-1/2}.$$

We let

$$\epsilon = 1 + \delta,$$

and since the wave-length in the guide in vacuum is given by

$$\lambda_{rg} = \lambda_{rs}[1 - (\lambda_{rs}/2b)^2]^{-1/2},$$

we have finally

$$\lambda_g/\lambda_{rg} = \left\{ 1 - \frac{\delta}{2}[1 - (\lambda_{rs}/2b)^2]^{-1} \right\},$$

treating δ as a quantity which is small compared to unity. ΔL is the shift measured in this investigation and the length of guide containing the gas is L . Let n be the number of the wave-lengths in the guide when evacuated. Then

$$n\lambda_{rg} = L.$$

We now admit gas to the guide but vary the position of the probe in such a manner as to follow a specific minimum in the standing wave pattern. Then

$$n\lambda_g = L - \Delta L,$$

and

$$\lambda_g/\lambda_{rg} = 1 - \Delta L/L.$$

Hence

$$\delta = \frac{2\Delta L}{L}[1 - (\lambda_{rs}/2b)^2]$$

which is Eq. (3).

When δ is so large that the approximation used in deriving (3) is not justified, we can always go back to the exact expression, namely

$$(\lambda_{rs}/\lambda_g)^2 + (\lambda_{rs}/2b)^2 = \epsilon.$$

Pressure-Volume Relation for Cylindrical Tubes with Elastomeric Walls: The Human Aorta

ALLEN L. KING

Dartmouth College, Hanover, New Hampshire

(Received December 19, 1945)

The general theory of the elasticity of elastomers as developed by James and Guth (reference 9) is extended and applied to cylindrical elastomeric tubes. Data obtained by Hallock and Benson (reference 4) on human aortas of several age groups are described rather well by the resulting equations. On the basis of this analysis an interpretation of the aging process for aortas is given.

DURING systole, blood from the left ventricle of the heart is forced into the aorta. From here it is distributed throughout the body by means of systemic arteries. With their elastic walls, the aorta and large systemic arteries apparently regulate the flow of blood to the capillary bed by smoothing the pulsating outflow from the heart—much as an electrical filter smooths out the pulsating d.c. from a rectifier circuit. With advancing age the arterial walls become less elastic and thicker as collagenous fibers increase in number and elastic tissue seemingly disappears. Lesions often occur, especially in old age, and senile degeneracy sets in. In spite of these several changes, the aorta and immediate systemic arteries continue to serve as a regulating system.^{1,2}

Evidently, elastic arteries have an important function in the cardiovascular system. At the same time, they are subject to an irreversible aging process which decreases their efficiency in smoothing out the pulsating blood flow. Many investigators have measured elastic characteristics of arterial walls in the hope that some insight would be gained regarding the nature of this aging process.

In 1880 Roy³ published an extensive report on the elastic properties of the arterial wall. He showed that animal tissues have thermoelastic properties comparable to those of caoutchouc and established the fact that static pressure-

volume curves for human aortas are normally S-shaped, especially in youth. In 1937 Hallock and Benson⁴ published measurements on segments of human aortas obtained post-mortem for several age groups and showed how the pressure-volume curve is changed in shape with advancing age. Observations on strips of aortas yield similar results.⁵ In all cases, normal healthy aortas have been found to possess non-linear characteristics and a decidedly rubber-like behavior. For convenience, therefore, in the following discussion an arterial wall will be considered replaced by a uniformly thick cylindrical *elastomeric* shell.

The aortic wall has been found to consist of three fairly well-differentiated layers.⁶ Its great extensibility is ascribed to the presence of elastic tissue in all three layers, but especially in the *tunica medialis*. The principal constituent of elastic tissue is an albuminoid called elastin. Interspersed with elastin are found other relatively inelastic substances. As a matter of fact, even these latter substances probably should be considered elastomeric. Because of their presence, the actual aortic wall has a greater thickness than the ideal one to be discussed below.

As mentioned above, an important factor generally recognized as influencing the rate of blood flow in the aorta is the distension of the wall. Elsewhere⁷ a general relation for the rate of flow of viscous fluids in cylindrical tubes with distensible walls has been given. In order to

¹ C. J. Wiggers, *Physiology in Health and Disease* (Lea and Febiger, Philadelphia, 1944), p. 555; H. D. Green, *Glasser's Medical Physics* (Year Book Publishers, Chicago, 1944), p. 208.

² E. V. Cowdry (ed.), *Arteriosclerosis* (The Macmillan Company, New York, 1933).

³ C. S. Roy, *Foster's J. Physiol.* 3, 125 (1880).

⁴ P. Hallock and I. C. Benson, *J. Clinical Invest.* 16, 595 (1937).

⁵ J. Krafka, *Arch. Pathology* 29, 303 (1940).

⁶ Reference 2, pp. 53-61.

⁷ A. L. King, *A. P. S. Bull.* 20, 8 (1945).

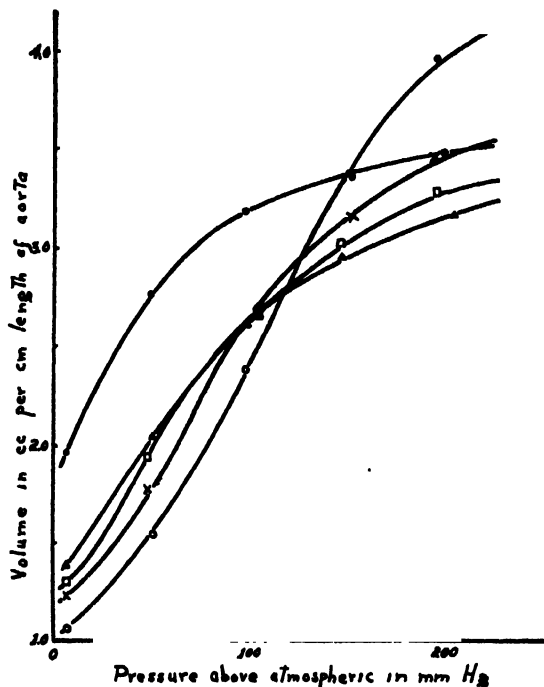


FIG. 1. Curves show Hallock and Benson's observed volumes per unit length of aorta as a function of applied fluid pressure. Theoretical values are plotted at five points for each age group: $\circ$ 20-24 yr., $\times$ 29-31 yr., $\square$ 36-42 yr., $\triangle$ 47-52 yr., $\bullet$ 71-78 yr.

obtain an explicit function for this rate, the variation of pressure with radius along the tube needs to be known. Although this variation should be a dynamic one, the static pressure-radius relation deduced below may be used to a first approximation.

During the preparation of this article, Rashevsky⁸ published in his journal an interesting treatment of the blood-flow problem. In his analysis he assumes the arterial wall to follow Hooke's law, although several well-known experimental results, such as those mentioned above, suggest that an entirely different approach should be made.

In this article a pressure-volume relation for cylindrical tubes with elastomeric walls is derived and applied to Hallock and Benson's data on human aortas.

THE PRESSURE-VOLUME RELATION

From the general theory of elastomers, the force exerted at right angles to a surface ys by a

⁸ N. Rashevsky, *Math. Biophys.* 7, 25, 35 (1945); see also O. Frank, *Zeits. f. Biol.* 71, 255 (1920).

mesh of molecular chains is given by^{9,10}

$$F_z = B_z y_0 z_0 \mathcal{L}^{-1}(k_z x), \quad (1)$$

where $y_0 z_0$ represents the undistorted area and $\mathcal{L}^{-1}(k_z x)$ is the inverse Langevin function of argument $k_z x$. k_z is a parameter which depends inversely on the maximum lengths of the molecular chains. If t represents the inverse Langevin function, then

$$\mathcal{L}(t) = \coth t - 1/t \quad (2)$$

is the Langevin function, where $t = F_z / B_z y_0 z_0$ and $\mathcal{L}(t) = k_z x$. The coefficient B_z has been assigned various values. For this discussion, however, the exact value does not enter. Nevertheless, it is of some interest to note that it increases directly with absolute temperature.

Consider a cylindrical tube of length z_0 with an elastomeric wall of thickness e_0 and internal

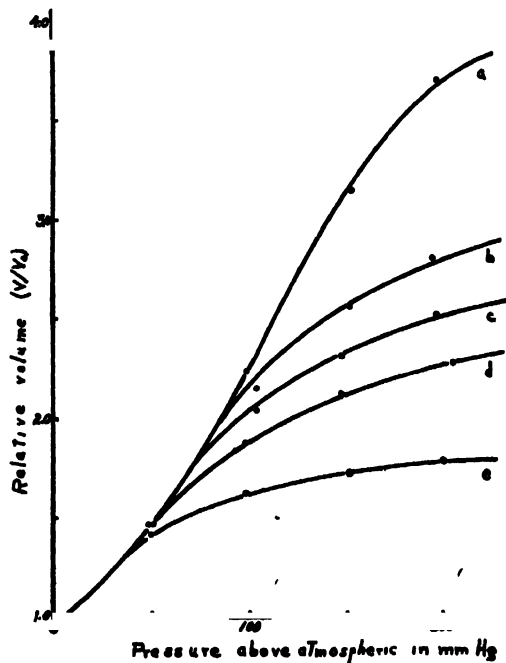


FIG. 2. Relative volumes of aortas in five age groups as a function of applied pressure. The curves are obtained from experimental data and the points from theory. (a) 20-24 yr., (b) 29-31 yr., (c) 36-42 yr., (d) 47-52 yr., (e) 71-78 yr.

⁹ F. T. Wall, *J. Chem. Phys.* 10, 485 (1942); 11, 527 (1943). H. M. James and E. Guth, *J. Chem. Phys.* 11, 455 (1943). See also A. V. Tobolsky and R. D. Andrews, *J. Chem. Phys.* 13, 3 (1945). E. Guth, "The Problem of the Elasticity of Rubber and Rubber-like Materials" in *Surface Chemistry*, F. R. Moulton (ed.) (A. A. S., Smithsonian Inst., Washington, D. C., 1943), pp. 103-127.

¹⁰ A. L. King, *Am. J. Phys.* 14, 30 (1946).

radius r_0 , oriented so that its axis lies along the z coordinate of a set of cylindrical polar axes. Initially, the pressure within and outside the tube is assumed to equal p_0 . For the moment, the weight of the tube is considered sufficiently small so that any distortion of the tube due to it may be neglected. As a further simplification, the thickness e_0 is assumed to be small compared with r_0 . Even for e_0 as much as 20 percent of r_0 this approximation appears justified, as suggested later in the application of this theory to human aortas.

Experimental¹¹ and theoretical¹² evidence indicates that the volume of a rubber-like material should remain constant as it is stretched up to 300 percent elongation. Thus, if the radius of the elastomeric wall changes to r due to an increased pressure within the tube, the thickness and length change to e and z , respectively, so that

$$2\pi r e z = 2\pi r_0 e_0 z_0. \quad (3)$$

For the present problem suppose e_0 and z_0 are multiplied by equal factors; then it follows that for Eq. (3) to be satisfied,

$$e = e_0(r_0/r)^{\frac{1}{2}} \quad \text{and} \quad z = z_0(r_0/r)^{\frac{1}{2}}. \quad (4)$$

In order to find the elastic forces within the wall, the randomly twisted and intermeshed molecular chains are replaced by a three-dimensional network of chains extending along the tube wall, around the tube wall, and radially between the outer and inner surfaces of the wall. These chains usually exert a tension on their colinear neighbors or on any free surface in which they end, since the most probable state for them is one of more complete coiling and random twisting. On the other hand, due to their sidewise thermal motions, the chains push outward on a free surface until the contractile, thermal, and any external forces thereon are in equilibrium. The thermal force may be considered to equal a uniform internal pressure P multiplied by the area of the free surface.

Since the molecular chains around the tube wall do not end on free surfaces, the internal

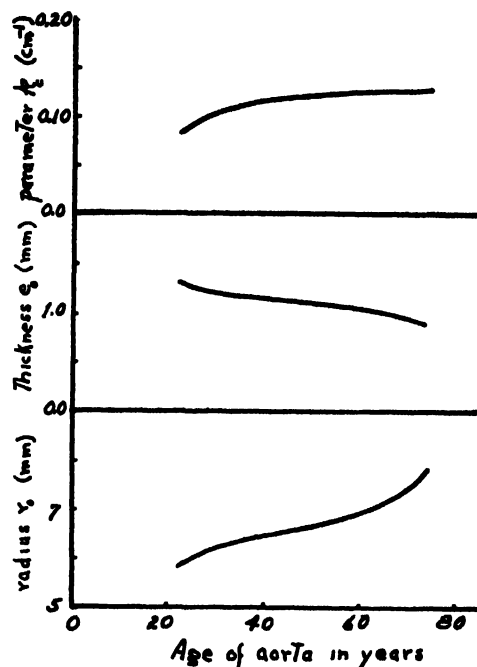


FIG. 3. Variation of radius, wall thickness, and a characteristic parameter with age of undistended aortas. The wall thickness here is an effective value. The parameter depends on the reciprocal length of molecular chains within the aortic wall.

peripheral force F_e comes out to be

$$F_e = e_0 z_0 B_e \mathcal{L}^{-1}(2\pi r k_e). \quad (5)$$

The longitudinal force is

$$F_l = 2\pi r_0 e_0 B_l \mathcal{L}^{-1}(k_l z) - 2\pi r e P; \quad (6)$$

and the radial force is

$$F_r = 2\pi r_0 e_0 B_r \mathcal{L}^{-1}(k_r e) - 2\pi r z P. \quad (7)$$

Pressure P may be found from the fact that F_r must equal $2\pi r z p_0$, since the outside surface of the wall is in equilibrium. Furthermore, the ends of the tube are supposed to be acted on by an external force Z which can become adjusted so that at all times $F_l + Z = 0$.

To obtain a relation between pressure p and radius r , a small increase δr in the radius is assumed to be accompanied by incremental changes δe and δz in thickness and length, respectively. Then, by the principle of virtual work:

$$2\pi r z (p - p_0) \cdot \delta r - Z \cdot \delta z = 2\pi F_e \cdot \delta r + F_r \cdot \delta e + F_l \cdot \delta z,$$

¹¹ W. L. Holt and A. T. McPherson, J. Research Nat. Bur. Stand. 17, 657 (1936).

¹² D. R. Elliot and S. A. Lippmann, J. App. Phys. 16, 50 (1945).

TABLE I. Constants for human aortas of various ages.

Age groups years	V_0 cc	r_0 mm	β	A mm Hg	e_0 mm	$e_0 r_0$ mm ²	k_r cm ⁻¹	d_0 cm
20-24	1.07	5.8	0.302	88	1.33	7.7	0.083	1.41
29-31	1.23	6.2	.415	74	1.20	7.45	.106	1.48
36-42	1.30	6.4	.462	70	1.17	7.5	.114	1.52
47-52	1.39	6.6	.510	65	1.12	7.4	.122	1.54
71-78	1.96	7.9	.640	44.5	0.92	7.3	.129	1.76

so that

$$p - p_0 = B_c \frac{e_0 z_0}{r z} \mathcal{E}^{-1}(2\pi r k_r) + p_0 (\delta e / \delta r). \quad (8)$$

But from Eq. (4), $\delta e = -\frac{1}{2}(e_0/r_0)(r_0/r)^{\frac{1}{2}} \cdot \delta r$ and letting $\beta = 2\pi r_0 k_r$, Eq. (8) reduces to

$$p - p_0 = (e_0/r_0) [B_c (r_0/r)^{\frac{1}{2}} \mathcal{E}^{-1}(\beta r/r_0) - \frac{1}{2} p_0 (r_0/r)^{\frac{1}{2}}]. \quad (9)$$

Since $p = p_0$ when $r = r_0$, evidently $p_0 = 2B_c \mathcal{E}^{-1}(\beta)$. Thus B_c may be eliminated from Eq. (9) to get the desired pressure-radius relation

$$p - p_0 = A \left[\frac{(r_0/r)^{\frac{1}{2}} \mathcal{E}^{-1}(\beta r/r_0)}{\mathcal{E}^{-1}(\beta)} - (r_0/r)^{\frac{1}{2}} \right], \quad (10)$$

where $A = e_0 p_0 / 2r_0$.

The initial volume within the tube is $V_0 = \pi r_0^2 z_0$ and the final volume is $V = \pi r^2 z$, so that $V/V_0 = (r/r_0)^2$; and the desired pressure-volume relation for cylindrical tubes with elastomeric walls is found to be

$$p - p_0 = A \left[\frac{(V_0/V)^{\frac{1}{2}} \mathcal{E}^{-1}\{\beta(V/V_0)^{\frac{1}{2}}\}}{\mathcal{E}^{-1}(\beta)} - (V_0/V)^{\frac{1}{2}} \right]. \quad (11)$$

Because k_r , which depends on the reciprocal of the maximum molecular chain lengths, cannot be determined readily, it is convenient to consider $\beta (= 2\pi r_0 k_r)$ an adjustable parameter of Eq. (11). Pressure p_0 is that of the surrounding medium; for the above analysis it is assumed constant. If the surrounding medium exerts a variable pressure, however, Eq. (11) can be modified accordingly.

Frequently the weight of the tube, and possibly other small extraneous factors, may cause a partial collapse of an elastomeric tube unless the pressure within is somewhat in excess of

that outside. This means the actual external pressure should be considered replaced by a somewhat greater effective pressure.

THE HUMAN AORTA

Perhaps the most widely quoted recent experimental data on human aortas are those of Hallock and Benson. Isolated segments, eight to ten centimeters long, of the descending aorta from just beyond the arch were obtained in the age groups 20-24, 29-31, 36-42, 47-52, and 71-78 years. As far as could be determined all specimens but those of the last group were healthy and normal. Those beyond seventy years showed some degree of coronary disease. The experimental curves are shown in Fig. 1.

To prevent collapse of the aortas an initial pressure of 10-cm saline solution (~ 7.5 mm Hg) was used in every case. The volumes per unit length for this initial pressure are given in Table I. Pressures were measured relative to atmosphere by means of a manometer. Since Hallock and Benson fail to state the prevailing atmospheric pressure for any of their observations, in the following discussion the standard value of 760 mm Hg is assumed to apply in all cases. But 7.5-mm Hg pressure are required to balance weight sag and other small factors, so that the effective external pressure p_0 is taken to be 768 mm Hg.

A better appreciation of the relative variation in pressure-volume characteristics with age may be obtained by plotting V/V_0 as a function of pressure difference as shown in Fig. 2. These curves are well represented by Eq. (11) of the above theoretical discussion. For convenience both β and A were adjusted so as to fit these curves. The resulting values are given in Table I. *The function is very sensitive to slight variations in β .*

Since V_0 represents the initial volume per unit length of aorta, the initial radius r_0 is given by $(V_0/\pi)^{\frac{1}{2}}$. Values of r_0 are listed in Table I and are plotted in Fig. 3. With these values of A and r_0 and $p_0 = 768$ mm Hg, the effective initial wall thicknesses e_0 are computed from $e_0 = 2r_0 A/p_0$ and also are plotted in Fig. 3. Notice that e_0 decreases with advancing age; probably this decrease is associated with the apparent disappearance of elastic tissue. On the other hand,

the product $e_0 r_0$, as seen in Table I, is constant, at least for the data under examination. The mean value is 7.5 mm^2 with an average deviation of only ± 0.1 . As dilation of an aorta proceeds with advancing age, the *effective* cross-sectional area of the tube wall remains nearly constant and equal to 47 mm^2 between the ages of 20 and 80 years. Perhaps this observation may be interpreted as follows: the materials within the vessel wall that render it elastic really do not disappear, but rather become more thinly distributed in the enlarged wall of the older aortas. More extensive experimental data, however, may show a real divergence from the above mean, especially at either end of the time scale, in one case due to new tissue growth and in the other to senile alterations.

External diameters d_0 of the undistended aortas approximately equal $2(r_0 + e_0)$. The values shown in Table I are in rather good agreement with measured values. For instance, Roy¹³ observed circumferences of 45 mm for a 22-year old and 52 and 55 mm for 71- and 76-year old aortas. These values correspond to diameters of 1.4 cm for the young aorta and 1.7 cm for the old aortas. Observed wall thicknesses, on the other hand, vary greatly, as illustrated by Krafka's data for *strips* of human aortas having no marked sclerotic lesions (Table II).⁵ These values, for the most part, are outside the limits of 1–2 mm mentioned by Krafka in an earlier paper.¹⁴ It should be recognized, however, that on cutting strips from an aorta, peripheral chains are shortened and made to end on free surfaces. The equilibrium is thereby altered and, perhaps, the wall then becomes thicker. In any case, as

TABLE II. Observed wall thicknesses of human aortas.

Age (years)	7	20	31	37	37	41	47	48
Thickness (mm)	1.8	2.6	2.6	2.9	1.76	3.18	2.34	2.46

suggested previously, e_0 should be expected to have values somewhat less than the actual wall thicknesses.

From $\beta = 2\pi r_0 k_e$ the parameter k_e is computed and entered in Table I. Its variation with age is shown in Fig. 3. Evidently k_e is an important age-dependent physical characteristic of aortas. Since it is defined as the molecular chain length reciprocal, clearly its increase with age corresponds to an effective decrease in distance between chain ends around the aorta. Perhaps a not unreasonable interpretation may be this: with advancing age, bonds between chains become established, thus making the effective chain lengths less and, consequently, producing a more rigid structure. In other words, arterial walls may be thought to lose their elasticity as a result of increased interchain bonding. Notice that chain lengths are reduced, on the average, by a factor of nearly two-thirds during the age interval from 20 to 80 years. The parameter β itself is a factor which represents the amount of coiling and twisting experienced by the peripheral molecular chains. Thus the distance between ends of a single chain around the aorta is only 0.302 of the maximum chain length for 20–24 year old specimens, whereas this ratio becomes 0.640 for 71–78 year old specimens.

The above analysis does not exclude the possibility of chains aggregated in strands. Perhaps it is such strands which produce the wavy dark streaks in sections of aortic wall that have been properly stained to show the presence of elastin.

¹³ C. S. Roy (reference 3), pp. 148 and 152–153.

¹⁴ J. Krafka, Am. J. Physiol. 125, 5 (1939).

Approximations to Planckian Distributions

PARRY MOON

Massachusetts Institute of Technology, Cambridge, Massachusetts

AND

DOMINA EBERLE SPENCER

Tufts College, Medford, Massachusetts

(Received January 22, 1946)

The use of Planck's equation in theoretical photometry and colorimetry has been hampered by the difficulty of integration. The following investigation attempts to eliminate this obstacle by replacing Planck's equation in the visible region by simpler expressions that are easily integrated. Two such approximations are employed—a three-term polynomial and the Wien function. Both allow analytic integrations to be performed with considerable ease and satisfactory accuracy. The method applies even in the weighted integrations of colorimetry: the standard weighting functions representing characteristics of the human eye are introduced in their analytic form and make possible the use of analytic methods in place of the unwieldy numerical approximations employed previously.

1. INTRODUCTION

THE importance of Planck's equation in photometry and colorimetry can hardly be overestimated. Unfortunately, however, the equation does not lend itself to easy manipulation. Ratios of Planck functions obstinately keep their form instead of simplifying as Wien functions do, and integrations are performed only with difficulty.

It seems doubtful if a simple approximation to the Planck equation is possible when the entire

wave-length scale is needed. But in photometry and colorimetry, only the visible region is considered, and in this region the radiation curves have very simple shapes which can be approximated by elementary functions.

The following pages present the results of an investigation of approximations to Planckian distributions. Both polynomials and Wien functions are employed. In either case, analytic integrations are easily performed and various kinds of mathematical manipulation are expedited.

TABLE I. Polynomial approximation to Planckian distributions. $J(\lambda) = K_0 + K_1\lambda + K_2\lambda^2$ (watt m⁻² micron⁻¹.)

T (°K)	K ₀	K ₁	K ₂
2000	2.15635 × 10 ⁴	- 9.7454 × 10 ⁴	11.1290 × 10 ⁴
2100	3.03280	-14.0909	16.5959
2200	3.95017	-19.0373	23.3151
2300	4.73990	-24.0395	30.9838
2400	5.15858	-28.2523	39.0340
2500	4.88558 × 10 ⁴	-30.7404 × 10 ⁴	46.6020 × 10 ⁴
2600	3.53338	-29.3414	52.5395
2700	6.65957	-23.0210	55.4300
2800	4.21592	- 9.6010	53.6250
2842	6.97760	- 1.4168	51.0905
2900	-11.57840	+12.9650	45.3400
3000	- 2.19248 × 10 ⁴	+ 4.6864 × 10 ⁴	2.8599 × 10 ⁴
3100	- 3.56774	9.4090	+ 0.1550
3200	- 5.32508	15.6716	+ 3.7805
3300	+ 7.49365	23.6450	- 9.1200
3400	-10.09439	33.4790	-16.0250
3500	+ 1.31403 × 10 ⁷	4.5300 × 10 ⁷	- 2.4642 × 10 ⁷
4000	3.48851	13.5166	- 9.5855
4500	4.04611	26.7200	-21.1150
5000	- 7.84826 × 10 ⁷	40.8458 × 10 ⁷	-34.9482 × 10 ⁷
5500	- 7.13490	50.9390	-47.4900
6000	- 2.10080	51.4390	-54.4858
6500	+ 2.89830	37.0600	-51.7500
7000	+ 2.71395 × 10 ⁸	3.5735 × 10 ⁸	- 3.5870 × 10 ⁸
20000	+ 3.09456 × 10 ¹⁰	- 8.4853 × 10 ¹⁰	+ 6.1053 × 10 ¹⁰

2. POLYNOMIAL APPROXIMATION

The simplicity of the plot of Planck's equation in the visible region suggests the possibility of approximating it by a few terms of a power series. At temperatures above 2500°K, a fairly satisfactory representation is obtained with two terms:

$$J(\lambda) = K_0 + K_2\lambda^2, \quad (1)$$

where K_0 and K_2 are constants and λ is the wave-length. At 2000°K, however, three terms are found to be necessary; and in the interests of standardization, three terms were adopted for approximating Planckian radiation at all temperatures.

The equation being approximated is

$$J(\lambda) = \frac{C_1}{\lambda^5} \frac{1}{\exp(C_2/\lambda T) - 1}, \quad (2)$$

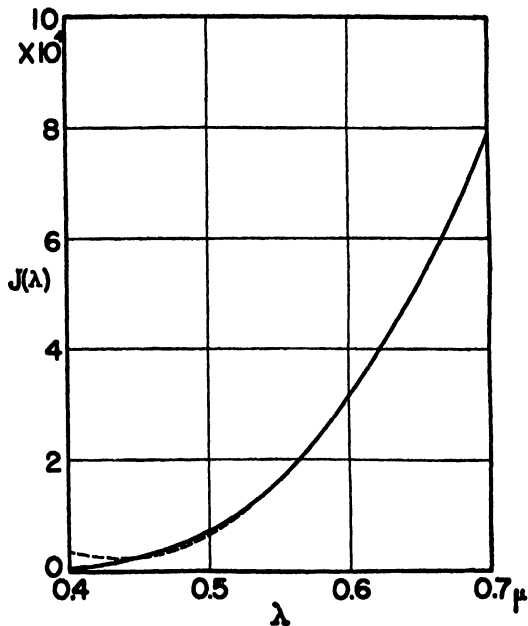


FIG. 1. Planckian radiation at 2000°K, showing the kind of fit that is obtained by the polynomial approximation at low temperatures. Ordinates are in watt m² micron⁻¹ at the radiator opening. — - Planckian radiation; - - - - three-term analytic approximation.

where¹ $C_1 = 3.6970 \times 10^8$, $C_2 = 14320$, λ = wavelength (micron), and $J(\lambda)$ = spectral pharosage (watt m⁻² micron⁻¹ at blackbody opening). The approximate equation is

$$J(\lambda) = K_0 + K_1\lambda + K_2\lambda^2. \quad (3)$$

If $J(\lambda)$ from Eq. (3) is made to coincide with $J(\lambda)$ from Eq. (2) at wave-lengths 0.45, 0.55, and 0.65 μ , the constants K_0 , K_1 , and K_2 are determined by the equations,

$$\begin{aligned} K_0 &= 17.875J(0.45) - 29.250J(0.55) \\ &\quad + 12.375J(0.65), \\ K_1 &= -60J(0.45) + 110J(0.55) \\ &\quad - 50J(0.65), \\ K_2 &= 50J(0.45) - 100J(0.55) + 50J(0.65). \end{aligned} \quad (4)$$

The constants of Eq. (3) are evaluated by means of Eq. (4) for a number of temperatures, and the results are listed in Table I. Values for other temperatures are easily computed by means of Eqs. (2) and (4). The discrepancies between

¹ The international temperature scale is used. See G. K. Burgess, *Bur. Stands. J. Res.* 1, 635 (1928); H. T. Wensel, *Bur. Stands. J. Res.* 22, 375 (1939); Parry Moon, *J. Math. and Phys.* 16, 133 (1937); M. I. T., *Elec. Eng. Dept., Contribution No. 131* (1938).

Planckian radiation and the three-term approximation may be visualized by considering Figs. 1 to 4. These graphs cover the range ordinarily employed in colorimetry and show that a reasonably good fit is obtained at all temperatures through most of the visible region. The discrepancies at the ends of the visible region do not prove troublesome in most applications. It is interesting to note that curves of such diverse shapes as Figs. 2 and 4 can be fitted by the same form of expression.

If a more accurate representation is desired, a seven-term polynomial may be used:

$$J(\lambda) = K_0 + K_1\lambda + K_2\lambda^2 + \dots + K_6\lambda^6. \quad (5)$$

The details in the evaluation of the constants are discussed elsewhere.² For example, take Planckian radiation at 2842°K and pass the approximate curve through the correct values at 0.40, 0.45, 0.50, $\dots$ 0.70 μ . The constants are found to be,

$$\begin{aligned} K_0 &= -1,185.36370, \\ K_1 &= +13,764.24041, \\ K_2 &= -63,762.05721, \\ K_3 &= +148,682.59032, \\ K_4 &= -182,496.89139, \\ K_5 &= +114,098.58438, \\ K_6 &= -28,888.87725. \end{aligned}$$

A comparison of the true values of $J(\lambda)$ and the values computed by Eq. (5) is shown in Table II. Values at 0.40, 0.45, etc., are of course correct, so only intermediate values are included in the table. It will be noted that the maximum error is only 1/50 of one percent.

TABLE II. Planckian radiation at 2842°K. $J(\lambda)$ in watt m⁻² micron⁻¹ for Planckian radiator. True values calculated by Planck's equation ($C_2 = 14320$), approximate values by seven-term polynomial, Eq. (5).

λ	$J(\lambda)$ (true)	$J(\lambda)$ (approx.)	Percent difference
0.425 μ	18.924×10^4	18.923×10^4	-0.005
0.475	37.799	37.791	-0.021
0.525	62.940	62.930	-0.016
0.575	92.018	92.003	-0.016
0.625	122.274	122.255	-0.016
0.675	151.245	151.219	-0.017

² Parry Moon and D. E. Spencer, "Polynomial representation of reflectance curves," *J. Opt. Soc. Am.* 35, 597 (1945). See particularly Table V.

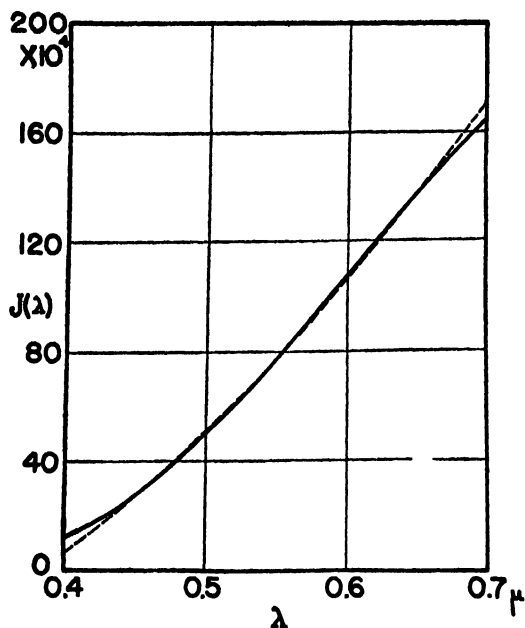


FIG. 2. Planckian radiation at 2842°K, showing the degree of approximation afforded by a three-term polynomial at a common operating temperature for incandescent lamps. — Planckian radiation; ---- three-term analytic approximation.

3. WIEN APPROXIMATION

Another method of approximating the Planckian distribution is by means of the Wien radiation equation. This approximation, while

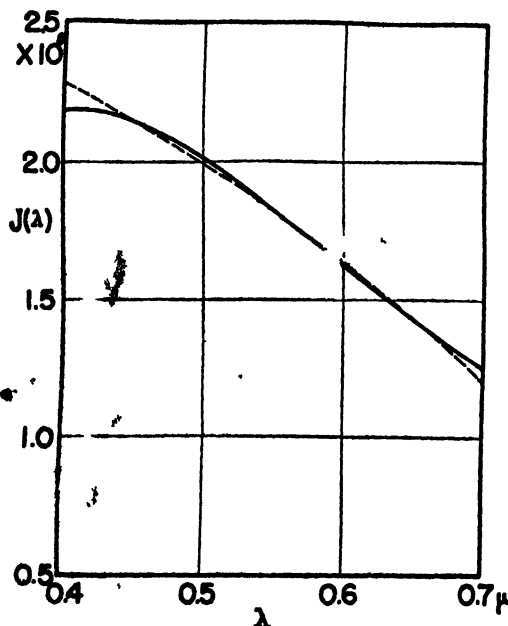


FIG. 3. Planckian radiation at 7000°K. Three-term polynomial approximation is shown dotted.

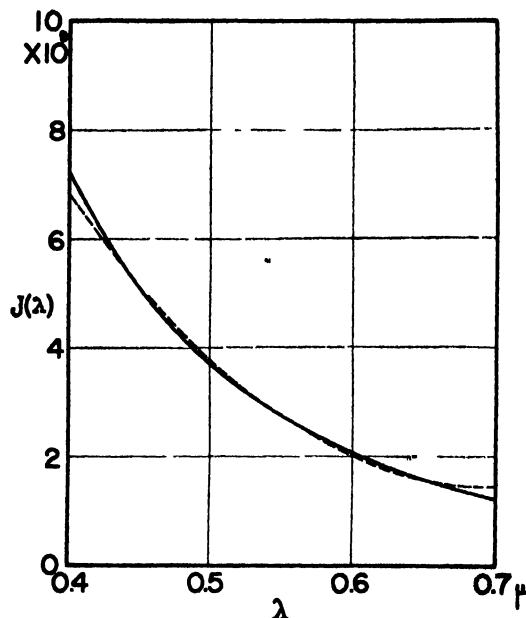


FIG. 4. Planckian radiation at 20,000°K. Three-term polynomial approximation is shown dotted.

not as simple as the three-term polynomial, allows comparatively easy integration. Since the Planck curve and the Wien curve are of different shapes and cannot be made to coincide, it is necessary to decide what aspect of the two will be made identical. Estey³ has used color temperature as the criterion by plotting the Wien points in the chromaticity diagram and obtaining shortest distances by Judd's method.⁴ Gage⁵ used the simple and definite criterion of making the two curves pass through the same points at 0.45 and 0.65μ.

We have chosen Gage's method but have altered both T and C_1 in order to obtain coincidence at 0.45 and 0.65μ. Previous work of this kind has been confined entirely to the change in quality, in accordance with the astonishing propensity of colorimetrists to ignore magnitude.

Planck's distribution is expressed in Eq. (2). The Wien distribution is represented by the equation

$$J(\lambda) = \frac{C_1'}{\lambda^5} \exp(-C_2/\lambda T'), \quad (6)$$

³ R. S. Estey, *J. Opt. Soc. Am.* **28**, 293 (1938).

⁴ D. B. Judd, *J. Opt. Soc. Am.* **26**, 421 (1936).

⁵ H. P. Gage, *J. Opt. Soc. Am.* **23**, 46 (1933); H. P. Gage and N. Macbeth, *Trans. I.E.S.* **31**, 995 (1936).

where C_2 has the same value as in Eq. (2) but C_1' and T' differ from the Planckian values. If the ordinates of the two curves are to be identical at wave-lengths λ_1 and λ_2 ($\lambda_2 > \lambda_1$),

$$\left. \begin{aligned} \frac{1}{C_1'} \exp(C_2/\lambda_1 T') &= \frac{1}{C_1} [\exp(C_2/\lambda_1 T) - 1], \\ \frac{1}{C_1'} \exp(C_2/\lambda_2 T') &= \frac{1}{C_1} [\exp(C_2/\lambda_2 T) - 1]. \end{aligned} \right\} \quad (7)$$

Dividing, and substituting

$$1/T' = 1/T + 1/\delta, \quad (8)$$

we obtain

$$\frac{1}{\delta} = \frac{\lambda_1 \lambda_2}{C_2(\lambda_2 - \lambda_1)} \ln \left[\frac{1 - \exp(-C_2/\lambda_1 T)}{1 - \exp(-C_2/\lambda_2 T)} \right]. \quad (9)$$

The quantity $1/\delta$ is the correction in reciprocal temperature that must be applied in going from the Planckian temperature to the corresponding temperature for the Wien distribution. This correction is always positive, corresponding to universal decrease in temperature from Planck to Wien. The advantage of Eq. (9) over those used by Gage and Estey is that the former is a correction term which is usually small and which therefore needs little accuracy in its computation.

TABLE III. Temperatures for Wien approximation.

T (Planck)	$1/\delta$ (mireds)	ΔT (°K)	T' (Wien)	ΔC_1	C_1' (Wien)
2000°K	0.00167	-0.0067	1999.9933	0.002 × 10 ⁷	3.6972 × 10 ⁶
2100	0.00281	0.0124	2099.9876	0.003	3.6973
2200	0.00442	0.0214	2199.9786	0.005	3.6975
2300	0.00697	0.0369	2299.9631	0.008	3.6978
2400	0.01035	0.0596	2399.9404	0.012	3.6982
2500	0.01490	0.0931	2499.9069	0.018	3.6988
2600	0.02085	0.1409	2599.8591	0.025	3.6995
2700	0.02844	0.2073	2699.7927	0.033	3.7003
2800	0.03791	0.2972	2799.7028	0.045	3.7015
2842	0.04252	0.3434	2841.6566	0.050	3.7020
2900	0.04953	0.4165	2899.5835	0.059	3.7029
3000	0.06354	0.5718	2999.4282	0.076	3.7046
3100	0.08018	0.7704	3099.2296	0.096	3.7066
3200	0.09967	1.0203	3198.9797	0.119	3.7089
3300	0.12222	1.3305	3298.6695	0.146	3.7116
3400	0.14803	1.7103	3398.2897	0.177	3.7147
3500	0.17725	2.1700	3497.8300	0.213	3.7183
4000	0.37919	6.0580	3993.9420	0.462	3.7432
4500	0.67991	13.7262	4486.2738	0.841	3.7811
5000	1.0778	26.8020	4973.1980	1.356	3.8326
5500	1.5631	46.8832	5453.1168	2.005	3.8975
6000	2.1216	75.4192	5924.5808	2.780	3.9750
6500	2.7377	113.6445	6386.3553	3.668	4.0638
7000	3.3962	162.5522	6837.4478	4.661	4.1631
20000	17.9979	5293.6610	14706.3391	45.350	8.2320

TABLE IV. Computed values of $J(\lambda)$ (watt m⁻² micron⁻¹).

λ (micron)	Planck	Wien	Percent differ- ence	Polynomial	Percent differ- ence
$T = 2842.0^\circ\text{K}, C_1 = 3.6970 \times 10^6, T' = 2841.66^\circ\text{K}, C_1' = 3.7020 \times 10^6$					
0.40	12.214 × 10 ⁴	12.212 × 10 ⁴	-0.017	6.302 × 10 ⁴	-48.3
0.45	27.475	27.475	0	27.475	0
0.50	49.711	49.716	+0.010	50.866	+2.3
0.55	77.159	77.170	+0.014	77.159	0
0.60	107.17	107.18	+0.010	105.65	-1.4
0.65	137.05	137.05	0	137.05	0
0.70	164.65	164.61	-0.024	170.65	+3.6
$T = 7000.0^\circ\text{K}, C_1 = 3.6970 \times 10^6, T' = 6837.46^\circ\text{K}, C_1' = 4.16312 \times 10^6$					
0.40	2.1830 × 10 ⁸	2.1638 × 10 ⁸	-0.88	2.2830 × 10 ⁸	+4.6
0.45	2.1483	2.1483	0	2.1483	0
0.50	2.0111	2.0204	+0.46	1.9959	-0.76
0.55	1.8254	1.8359	+0.57	1.8254	0
0.60	1.6254	1.6320	+0.41	1.6370	+0.71
0.65	1.4306	1.4306	0	1.4306	0
0.70	1.2508	1.2432	-0.61	1.2065	-3.5
$T = 20,000^\circ\text{K}, C_1 = 3.6970 \times 10^6, T' = 14,706.3^\circ\text{K}, C_1' = 8.23200 \times 10^6$					
0.40	7.2360 × 10 ⁸	7.0468 × 10 ⁸	-2.6	6.7731 × 10 ⁸	-6.4
0.45	5.1251	5.1251	0	5.1251	0
0.50	3.7120	3.7574	+1.2	3.7827	+1.9
0.55	2.7451	2.7847	+1.4	2.7451	0
0.60	2.0689	2.0890	+0.97	2.0134	-2.7
0.65	1.5862	1.5862	0	1.5862	0
0.70	1.2350	1.2187	-1.3	1.4651	+18.6

This advantage is particularly noticeable at the lower temperatures.

Equation (9) is exact. Usually, however, the exponentials are small in comparison with unity and a simpler expression is possible. Since

$$\ln(1-x) = -[x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \dots],$$

Eq. (9) may be written,

$$\frac{1}{\delta} \cong \frac{\lambda_1 \lambda_2}{C_2(\lambda_2 - \lambda_1)} \times [\exp(-C_2/\lambda_2 T) - \exp(-C_2/\lambda_1 T)]. \quad (10)$$

If $\lambda_1 = 0.450\mu$, $\lambda_2 = 0.650\mu$, and $C_2 = 14,320$, Eq. (9) becomes

$$\frac{1}{\delta} = 1.0213 \times 10^{-4} \ln \left[\frac{1 - \exp(-31,822/T)}{1 - \exp(-22,031/T)} \right] \quad (9a)$$

and Eq. (10) is

$$\frac{1}{\delta} \cong 1.0213 \times 10^{-4} [\exp(-22,031/T) - \exp(-31,822/T)]. \quad (10a)$$

Next the constant C_1' is evaluated. From Eqs. (2) and (6),

$$C_1' = C_1 \frac{\exp(C_2/\lambda_1 T')}{[\exp(C_2/\lambda_1 T) - 1]}. \quad (11)$$

Let

$$C_1' = C_1 + \Delta C_1.$$

Then the increment in C_1 , to be used in changing from Planck to Wien radiation, is

$$\Delta C_1 = C_1' - C_1$$

or

$$\Delta C_1 = C_1$$

$$\times \left[\frac{\exp(C_2/\lambda_1 \delta) - 1 + \exp(-C_2/\lambda_1 T')}{1 - \exp(-C_2/\lambda_1 T')} \right]. \quad (12)$$

Equation (12) is exact. At temperatures below approximately 3500°K it may be approximated by

$$\Delta C_1 \cong C_1 \left[\frac{C_2}{\lambda_1 \delta} + \exp(-C_2/\lambda_1 T') \right]. \quad (12a)$$

Below 2800°K, the exponential may be neglected, resulting in the simple equation,

$$\Delta C_1 \cong \frac{C_1 C_2}{\lambda_1 \delta}. \quad (12b)$$

If $\lambda_1 = 0.45\mu$, $C_1 = 3.6970 \times 10^8$, $C_2 = 14,320$, Eq. (12) becomes

$$\Delta C_1 = 3.6970 \times 10^8 \times \left[\frac{\exp(31822 \cdot 1/\delta) - 1 + \exp(-31822/T')}{1 - \exp(-31822/T')} \right] \quad (13)$$

and Eq. (12a) reduces to

$$\Delta C_1 \cong 3.6970 \times 10^8 [31822 \cdot (1/\delta) + \exp(-31822/T')]. \quad (13a)$$

Also Eq. (12b) becomes

$$\Delta C_1 \cong 11.765 \times 10^{12} (1/\delta). \quad (13b)$$

Values of $1/\delta$ and ΔC_1 are listed in Table III. Also given are values of T' , $\Delta T'$, and C_1' . A comparison of computed values of $J(\lambda)$ is shown in Table IV. The values calculated by Planck's equation are listed in the second column, and the Wien values are in the third column. Computations made by use of the three-term polynomial approximation are also given (fifth column). Table IV shows that at 2842°K, the maximum error caused by using the Wien equation instead of the Planck equation is only 1/40 percent.

TABLE V. Radiant pharosage in visible spectrum. D_r (watt m^{-2}) for Planckian radiator.

	7 = 2842°K	7000	20,000
Planck	2.4290×10^4		
Wien	2.4290×10^4	5.3933×10^7	9.6566×10^8
2-term polynomial	2.4264×10^4	5.3929×10^7	
(2-term-Wien)/Wien	0.0011	0.00007	
3-term polynomial	2.4244×10^4	5.3955×10^7	9.6090×10^8
(3-term-Wien)/Wien	0.0019	0.00041	0.0049
$D_r(\text{visible region})/D_r(\text{total})$	0.0652	0.3935	0.0106

For higher temperatures, the discrepancies between the Wien and Planck distributions are more pronounced. For 20,000°K, differences of one or two percent are encountered at some wave-lengths between the true curve and the correlated Wien curve ($T' = 14,706$, $C_1' = 8.232 \times 10^8$). The Wien distribution is always too peaked, giving values that are too high in the center and too low at the ends of the visible spectrum. The three-term polynomial representation is inferior in accuracy to the Wien distribution at low temperatures; but at temperatures above 7000°K, the two are not widely different.

4. INTEGRATION WITHOUT WEIGHTING FUNCTION

It is sometimes necessary to obtain the radiant power in a specified wave-length band. For instance, the portion of the radiant power of a Planckian radiator may be required for the visible region or for a portion of the visible region between the wave-lengths λ_1 and λ_2 . The Planck equation then proves to be inconvenient since it can be integrated only by expansion in an infinite series of exponentials. A numerical method of integration was developed by Holladay,⁶ based on the tables of Frehafer and Snow.⁷ The work differs completely from the present paper in that analytic methods were not used.

The integral to be evaluated is

$$D_r = \int_{\lambda_1}^{\lambda_2} J(\lambda) d\lambda, \quad (14)$$

where D_r = radiant pharosage (watt m^{-2}) between the wave-lengths λ_1 and λ_2 , and $\lambda_2 > \lambda_1$ (micron).

⁶ L. L. Holladay, J. Opt. Soc. Am. 17, 329 (1928).

⁷ M. K. Frehafer and C. S. Snow, Miscellaneous Publications of the Bureau of Standards, No. 56 (1925).

If a three-term polynomial approximation is employed,

$$J(\lambda) = K_0 + K_1\lambda + K_2\lambda^2, \quad (3)$$

and the integration is very simple:

$$\begin{aligned} D_r &= \int_{\lambda_1}^{\lambda_2} (K_0 + K_1\lambda + K_2\lambda^2) d\lambda, \\ &= K_0(\lambda_2 - \lambda_1) + \frac{K_1}{2}(\lambda_2^2 - \lambda_1^2) \\ &\quad + \frac{K_2}{3}(\lambda_2^3 - \lambda_1^3). \end{aligned} \quad (15)$$

As an example of the use of Eq. (15), Table V was computed. It gives the radiant pharosage in the wave-length band from 0.40 to 0.70 μ for Planckian radiators operating at several temperatures. The values calculated by means of Eq. (15) are compared with the correct values obtained by series integration of Planck's equation. The discrepancies are seen to be less than a half percent in all cases.

Now consider the integration of the generalized Wien equation:

$$J(\lambda) = C_1\lambda^{-5} \exp(-nC_2/\lambda T). \quad (16)$$

Equation (14) becomes

$$D_r = \int_{\lambda_1}^{\lambda_2} J(\lambda) d\lambda = C_1 \int_{\lambda_1}^{\lambda_2} \lambda^{-5} \exp(-nC_2/\lambda T) d\lambda$$

or

$$\begin{aligned} D_r &= C_1 \left(\frac{T}{nC_2} \right)^4 \left\{ \exp(-nC_2/\lambda_2 T) \cdot \left[\left(\frac{nC_2}{\lambda_2 T} \right)^3 + 3 \left(\frac{nC_2}{\lambda_2 T} \right)^2 + 6 \left(\frac{nC_2}{\lambda_2 T} \right) + 6 \right] \right. \\ &\quad \left. - \exp(-nC_2/\lambda_1 T) \cdot \left[\left(\frac{nC_2}{\lambda_1 T} \right)^3 + 3 \left(\frac{nC_2}{\lambda_1 T} \right)^2 + 6 \left(\frac{nC_2}{\lambda_1 T} \right) + 6 \right] \right\}. \end{aligned} \quad (17)$$

For Planckian radiation, a series expansion is employed:

$$\begin{aligned} J(\lambda) &= \frac{C_1}{\lambda^5} \frac{1}{\exp(C_2/\lambda T) - 1}, \\ &= \frac{C_1}{\lambda^5} [\exp(-C_2/\lambda T) + \exp(-2C_2/\lambda T) + \dots], \end{aligned} \quad (18)$$

and the integration of each term is obtained by use of Eq. (17) which becomes

$$\begin{aligned} D_r &= 0.87918(T/n)^4 \left\{ \exp(-14320n/\lambda_2 T) \cdot \left[\left(\frac{14320n}{\lambda_2 T} \right)^3 + 3 \left(\frac{14320n}{\lambda_2 T} \right)^2 + 6 \left(\frac{14320n}{\lambda_2 T} \right) + 6 \right] \right. \\ &\quad \left. - \exp(-14320n/\lambda_1 T) \cdot \left[\left(\frac{14320n}{\lambda_1 T} \right)^3 + 3 \left(\frac{14320n}{\lambda_1 T} \right)^2 + 6 \left(\frac{14320n}{\lambda_1 T} \right) + 6 \right] \right\}. \end{aligned} \quad (17a)$$

Evidently Eq. (18) leads to great complication, particularly at high temperatures. Simplification is effected by using the Wien equation ($n=1$) with temperature T' and constant C_1' . Then Eq. (17) becomes

$$\begin{aligned} D_r &= C_1'(T'/14320)^4 \left\{ \exp(-14320/\lambda_2 T') \cdot \left[\left(\frac{14320}{\lambda_2 T'} \right)^3 + 3 \left(\frac{14320}{\lambda_2 T'} \right)^2 + 6 \left(\frac{14320}{\lambda_2 T'} \right) + 6 \right] \right. \\ &\quad \left. - \exp(-14320/\lambda_1 T') \cdot \left[\left(\frac{14320}{\lambda_1 T'} \right)^3 + 3 \left(\frac{14320}{\lambda_1 T'} \right)^2 + 6 \left(\frac{14320}{\lambda_1 T'} \right) + 6 \right] \right\}. \end{aligned} \quad (17b)$$

TABLE VI. Coefficients for integration of polynomial approximation, light sources.

	X/K_m		Y/K_m		Z/K_m	
0	010666	57240	010685	35175	010677	25053
1	060921	26156	059773	82336	048473	18484
2	035226	72563	033616	34585	022066	86373
3	020585	29574	019006	76935	010073	48896
4	012138	18444	010804	00986	046112	93726
5	072128	04822	061741	99087	021167	64643
6	043147	45424	035472	77575	097438	95977
7	025962	45164	020489	36516	044978	64475
8	015703	36300	011898	18699	020820	78752
9	095428	23493	069462	88390	096551	46038
10	067101	94957	040770	52600	044992	86102
0	010666	57240	010685	35175	010677	25053
-1	018945	52333	019203	57729	023583	64653
-2	034207	93241	034696	72926	052233	94835
-3	062918	49167	063023	83522	1.1600	63798
-4	1.1809	92005	1.1508	72075	2.5834	18571
-5	2.2652	57902	2.1127	61624	5.7688	38040
-6	4.4431	71306	3.8991	48094	12.916	93227
-7	8.9117	26706	7.2340	01867	29.000	44752
-8	18.261	13996	13.491	81539	65.286	16073
-9	38.177	40910	25.295	17554	147.36	87925
-10	81.275	83667	47.673	01938	333.54	52442

Note.—Superscripts refer to the number of zeros after the decimal point. Thus 07212804822 = 0.007212804822.

Values of D_r computed by means of Eq. (17b) for $\lambda_1 = 0.40$ and $\lambda_2 = 0.70\mu$ are listed in Table V. At 2842°K, the value obtained from Eq. (17b) agrees with the true Planckian value, Eqs. (16a) and (17), in the fifth digit. The Planckian results are

$$\begin{aligned} n=1, \quad D_r &= 2.4283 \times 10^5; \\ n=2, \quad D_r &= 0.0007 \times 10^5; \\ n=3, \quad D_r &= \text{negligible.} \end{aligned}$$

TABLE VII. Three-term polynomial approximations.

	$T = 2000^\circ\text{K}$	2842	3500	7000	20,000
X , Planck	2812.4	97,871	49221 ¹	18413 ³	28581 ⁴
Polynomial approximation	2834.5	96,917	49007 ¹	18490 ³	28219 ⁴
Percent difference	0.79	+0.98	-0.44	+0.42	-1.3
Y , Planck	2203.0	89,030	47543 ¹	19073 ³	28727 ⁴
Polynomial approximation	2201.0	88,721	47531 ¹	19077 ³	28705 ⁴
Percent difference	-0.091	-0.34	-0.025	+0.021	-0.077
Z , Planck	326.03	31,808	25058 ¹	22738 ³	54090 ⁴
Polynomial approximation	343.04	31,371	24905 ¹	22794 ³	53830 ⁴
Percent difference	+4.5	-1.4	-0.61	+0.25	-0.48
x , Planck	0.52653	0.44749	0.40404	0.30574	0.25656
Polynomial	0.52700	0.44660	0.40354	0.30632	0.25479
Difference	+4.7 × 10 ⁻⁴	-8.9 × 10 ⁻⁴	-5.0 × 10 ⁻⁴	+5.8 × 10 ⁻⁴	-17.7 × 10 ⁻⁴
y , Planck	0.41243	0.40707	0.39027	0.31671	0.25788
Polynomial	0.40922	0.40884	0.39138	0.31605	0.25918
Difference	-32.1 × 10 ⁻⁴	+17.7 × 10 ⁻⁴	+11.1 × 10 ⁻⁴	-6.6 × 10 ⁻⁴	+13.0 × 10 ⁻⁴

Note.—Superscripts denote the number of zeros before the decimal point. For instance, 49221¹ = 492,210.

¹ Parry Moon and D. E. Spencer, "Analytical representation of trichromatic data," J. Opt. Soc. Am. 35, 399 (1945); and ² Analytic expressions in photometry and colorimetry," to appear in J. Math. and Phys.

The summation of these values gives the true result, $D_r = 2.4290 \times 10^5$ watt m⁻². Note that the results obtained by using our corresponding T' and C_1' are correct to five figures, while the common practice of assuming that $T' = T$ and $C_1' = C_1$ gives a result that is low by 7 in the fifth figure. At higher temperatures the exact computation becomes very laborious and the approximate methods are correspondingly advantageous.

5. INTEGRATION WITH WEIGHTING FUNCTIONS

Even more important than the integrations described in the preceding section are the weighted integrations that occur so frequently in colorimetry. If the weighting functions are $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$, then

$$X = \int_0^\infty \bar{x}(\lambda) J(\lambda) d\lambda,$$

$$Y = \int_0^\infty \bar{y}(\lambda) J(\lambda) d\lambda, \quad (19)$$

$$Z = \int_0^\infty \bar{z}(\lambda) J(\lambda) d\lambda.$$

If $J(\lambda)$ is expressed as a polynomial,

$$J(\lambda) = \sum_m K_m \lambda^m \quad (20)$$

and the weighting functions are given by their analytic representations,⁸ then the integrals are

evaluated analytically by use of the formulas,

$$(L/K)_m = A \frac{\Gamma(p-m-1)}{q^{p-m-1}},$$

$$(L/K)_{m+1} = (L/K)_m \frac{q}{p-m-2}, \quad (21)$$

$$(L/K)_{m-1} = (L/K)_m \frac{p-m-1}{q},$$

where L represents X , Y , or Z and where p and q are the constants given previously.⁸ These equations were used in the calculation of the constants of Table VI.

To obtain the trichromatic specification of any radiation whose spectral distribution can be approximated by a polynomial, one multiplies the polynomial coefficients K_m by the corresponding values of Table VI and adds:

$$L = \sum_m (L/K_m) K_m. \quad (22)$$

For Planckian radiation, the values of K_m for a three-term polynomial approximation may be employed. Examples are given in Table VII. Generally the error in X , Y , or Z caused by the polynomial approximation is less than one percent. The homogeneous coordinates x and y are also given in the table and the discrepancy caused by polynomial approximation is usually less than 20×10^{-4} .

If greater accuracy is desired, a seven-term polynomial may be employed. Table VIII shows that at 2842°K the trichromatic coordinates X , Y , and Z are in error because of polynomial approximation by less than 1/50 percent, while x and y are within 0.1×10^{-4} of the Planckian value obtained by the analytic method.

When $J(\lambda)$ is expressed in terms of a corresponding Wien distribution

$$J(\lambda) = \frac{C_1'}{\lambda^5} \exp(-C_2/\lambda T'), \quad (23)$$

the trichromatic coefficients of the radiation are found by evaluating the expression⁸

$$L = \frac{A C_1' T' (p+4)}{(q + C_2/T')^{p+4}} \quad (24)$$

TABLE VIII. Seven-term polynomial approximation at 2842°K.

X , Planck	97,871
Polynomial approximation	97,856
Percent difference	-0.015
Y , Planck	89,030
Polynomial approximation	89,016
Percent difference	-0.017
Z , Planck	31,808
Polynomial approximation	31,802
Percent difference	-0.019
x , Planck	0.44749
Polynomial approximation	0.44750
Difference	$+0.1 \times 10^{-4}$
y , Planck	0.40707
Polynomial approximation	0.40707
Difference	0.0×10^{-4}

for Z , X_B , X_C , Y_A , Y_D , and X_A and by using the relations

$$\begin{aligned} X &= X_A + X_C - X_B, \\ Y &= Y_A - Y_D. \end{aligned} \quad (25)$$

The corresponding values of C_1' and T' for Planckian distributions, Table IV, were used in evaluating X , Y , and Z . The results are shown in Table IX. It will be noted that X , Y , and Z are always within one percent of the values obtained with Planckian distributions. As would be expected, the errors are only a few thousandths of one percent at low temperatures. The homogeneous coordinates are accurate to five decimal places at low temperatures, and at 20,000°K are still accurate enough for practical purposes.

The total lamprosimy $\bar{y}_t$ (luminous efficacy) is also computed very easily by the methods of this paper. By definition,

$$\bar{y}_t = \int_0^\infty \bar{y}(\lambda) J(\lambda) d\lambda / \int_0^\infty J(\lambda) d\lambda. \quad (26)$$

Multiplication⁸ by 648 gives lumens per watt. The numerator is obtained from Eq. (21) or (23), while the denominator is obtained from the Stefan-Boltzmann relation,

$$\int_0^\infty J(\lambda) d\lambda = 5.709 \times 10^{-8} T^4 \text{ watt/m}^2. \quad (27)$$

When Planckian radiation is transmitted through a selective medium or reflected from a surface, the additional weighting function $\tau(\lambda)$ or

TABLE IX. Comparison of Planck and Wien equations, where Wien temperature and C_1' are adjusted in accordance with Eqs. (9) and (12).

	$T = 2000^\circ\text{K}$	2842	3500	7000	20,000
X, Planck	2812.4	97,871	49221 ¹	18413 ²	28581 ⁴
Wien	2812.5	97,879	49237 ¹	18467 ²	28741 ⁴
Percent difference	+0.0036	+0.0082	+0.037	+0.29	+0.56
Y, Planck	2203.0	89,030	47543 ¹	19073 ²	28727 ⁴
Wien	2203.0	89,040	47563 ¹	19165 ²	29079 ⁴
Percent difference	0	+0.011	+0.042	+0.48	+0.87
Z, Planck	326.03	31,808	25058 ¹	22738 ²	54090 ⁴
Wien	326.03	31,808	25059 ¹	22737 ²	54013 ⁴
Percent difference	0	0	+0.0040	-0.0044	-0.14
x, Planck	0.52653	0.44749	0.40404	0.30574	0.25656
Wien	0.52653	0.44749	0.40405	0.30590	0.25700
Difference	0.0×10^{-4}	0.0×10^{-4}	$+0.1 \times 10^{-4}$	$+1.6 \times 10^{-4}$	$+4.4 \times 10^{-4}$
y, Planck	0.41243	0.40707	0.39027	0.31671	0.25788
Wien	0.41243	0.40708	0.39031	0.31746	0.26002
Difference	0.0×10^{-4}	$+0.1 \times 10^{-4}$	$+0.4 \times 10^{-4}$	-7.5×10^{-4}	$+21.4 \times 10^{-4}$

$\rho(\lambda)$ is introduced. If the spectral reflectance or transmittance curves are represented also by polynomials, the integrations can be handled as in the preceding paragraphs. This subject is treated in a previous publication.² The total luminous reflectance is

$$\rho_l = \int_0^\infty \rho(\lambda) \bar{y}(\lambda) J(\lambda) d\lambda / \int_0^\infty \bar{y}(\lambda) J(\lambda) d\lambda, \quad (28)$$

where numerator and denominator are evaluated analytically. A similar expression applies to the total luminous transmittance τ_l .

6. CONCLUSIONS

An investigation has been made of the possibility of simplifying computations dealing with Planckian radiation. It is found that Planckian distributions can be approximated in the visible

region by three-term polynomials. Integrations can be made analytically with surprising ease and the results are generally correct to within one percent. Values of the coordinates x and y in the chromaticity diagram obtained by this method are in error ordinarily by less than 20×10^{-4} .

If greater accuracy is desired, it can be obtained by the same methods, using a larger number of terms in the polynomial. Data are given for a seven-term approximation which gives accuracy to five figures.

The Planckian distribution can be approximated also by a Wien distribution. The temperature is altered and a modified constant C_1 is used. Equations and tables are given so that the corresponding temperatures and values of C_1 can be obtained. The Wien distribution is more accurate than the three-term polynomial but is more trouble to compute.

Calendar of Meetings

June

- 20-22 American Association of Physics Teachers, St. Louis, Missouri
- 20-22 American Physical Society, Chicago, Illinois
- 20-23 Society for the Promotion of Engineering Education, St. Louis, Missouri

- 24-28 American Institute of Electrical Engineers, Detroit, Michigan
- 24-28 American Society for Testing Materials, Buffalo, New York

July

- 12-13 American Physical Society, Berkeley, California
- 17-19 American Society of Civil Engineers, Spokane, Washington

August

- 14-15 Institute of Aeronautical Sciences, Los Angeles, California

- 19-20 Mathematical Association of America, Ithaca, New York
- 20-23 American Mathematical Society, Ithaca, New York

September

- 9-13 American Chemical Society, Chicago, Illinois
- 16-20 The Instrument Society of America, Pittsburgh, Pennsylvania
- 18-21 Illuminating Engineering Society, Quebec, Canada
- 19-21 American Physical Society, New York, New York
- 30-October 3 American Society of Mechanical Engineers, Boston, Massachusetts

October

- 3-5 National Electronics Conference, Inc., Chicago, Illinois
- 9-11 American Society of Civil Engineers, Kansas City, Missouri
- 25 American Mathematical Society, New York, New York

Approximate Calculation of the Range of a Winged Jet Projectile

M. ZBIGMEW KRZYWOBLOCKI

Palo Alto, California

(Received January 2, 1946)

The purpose of this note is to discuss the possibility of launching winged jet bombs or mail-rockets on very long flights of perhaps 3000 miles, but the problem is considered from one standpoint only, i.e., that of mass ratio. There will be no discussion of the directional and stability instruments of a long distance winged jet projectile and the real mechanics of flight of such a unit, nor on various problems connected with winds, accuracy, etc. The problem will be treated in simplified form from the point of view of mass ratio at the beginning and at the end of flight. The simplified equations of 3 periods of flight are given: climbing, horizontal flight, and gliding. The simplifications introduced present a very rough approximation to the actual problem, but exact solution is not possible because of mathematical difficulties. A short example at the end of this note shows that the required mass ratio for a flight of 2000 or 3000 miles is high.

1. CLIMBING

ASSUME the flight path of a winged jet projectile consuming external air to be a straight line inclined to the level at an angle ϕ . The initial velocity is equal to zero. The reactional force of the exhaust gases is

$$F = m_f[c + \alpha(c - v)], \quad (1)$$

where

F , reactional force of the outflowing gases.

c , velocity of the outflowing gases with respect to the nozzle.

m_f , mass of the propulsive material (fuel).

α , air/fuel ratio by weight or by mass.

v , velocity of flight.

Substitute the following expression

$$m_f = (1/g)kW, \quad (2)$$

where

W , instantaneous weight of the jet bomb.

k , coefficient.

g , acceleration of gravity at sea level.

Neglect as a first approximation the influence of height on the value of the acceleration of gravity.

Thus

$$F = (1/g)kW[c + \alpha(c - v)]. \quad (3)$$

The rate of fuel injection is

$$dW = -kWdt. \quad (4)$$

From (4) with the initial condition, i.e., for $t=0$, $W=W_0$, one obtains

$$W = W_0 \exp \left[- \int_0^t k dt \right]. \quad (5)$$

Express the change in density of the air by the Hohmann's formula¹

$$\frac{\rho}{\rho_0} = \left[1 - \frac{h}{400,000} \right]^{49}, \quad (6)$$

where

ρ , density at the given height.

ρ_0 , density at sea level.

h , height.

The lift at any height may be given by the formula

$$L = W \cos \phi = (\rho/2)c_L S v^2, \quad (7)$$

where

L , lift.

c_L , lift coefficient.

S , wing area.

At sea level one has

$$W_0 = (\rho_0/2)c_L S v_0^2, \quad (8)$$

where v_0 denotes the velocity of horizontal flight at sea level. In a motion with subsonic velocity it is permissible to assume that the coefficient c_L is constant. Thus one obtains:

$$\left(\frac{W}{W_0} \right) \cos \phi = \frac{L}{W_0} = \left(\frac{v^2}{v_0^2} \right) \left[1 - \frac{h}{400,000} \right]^{49}. \quad (9)$$

Similarly the drag may be expressed in the form

$$D = \frac{\rho}{2} c_D S v^2 = \epsilon L, \quad (10)$$

¹Eugen Sänger, *Raketen flugtechnik* (Verlag von R. Oldenbourg, München and Berlin, 1933).

where

c_D , drag coefficient assumed to be constant in subsonic flight.

ϵ , c_D/c_L .

From the equation of equilibrium of forces in the direction of motion one obtains

$$\frac{F}{W} = \sin \phi + \epsilon \cos \phi + \left(\frac{1}{g}\right) \left(\frac{dv}{dt}\right). \quad (11)$$

Assume as a first, very rough approximation that W is a constant and equal to the mean of the values of the weight of the rocket at the beginning and end of the climbing period:

$$W = \frac{W_0 + W_1}{2}, \quad \frac{W}{W_0} = \frac{W_0 + W_1}{2W_0} = k_1 = \text{const.}, \quad (12)$$

where

W_1 , weight of the jet projectile at the end of the climbing period.

From (9) one obtains

$$v \frac{ds}{dt} = v_0 (k_1 \cos \phi)^{\frac{1}{2}} \left[1 - \frac{s \sin \phi}{400,000} \right]^{-24.5}, \quad (13)$$

where

s , distance of flight measured along the path.

t , time.

h , $s \sin \phi$.

Integration of (13) between the limits 0 and t with the initial condition $t=0$ for $s=0$ gives

$$t \left[\frac{15,689}{v_0 (k_1 \cos \phi)^{\frac{1}{2}} \sin \phi} \right] \times \left[1 - \left(1 - \frac{s \sin \phi}{400,000} \right)^{25.5} \right], \quad (14)$$

$$\frac{W_1}{W_0} = 2k_1 - 1 = \exp \left\{ - \int_0^t \frac{g [\sin \phi + \epsilon \cos \phi + AB^{-1.96}]}{c(\alpha+1) - CB^{-0.96}} dt \right\}, \quad (20)$$

$$C = av_0 (k_1 \cos \phi)^{\frac{1}{2}} = \alpha II; \quad (20a)$$

$$\frac{W_1}{W_0} = \exp \left\{ - \int_0^v \frac{g [\sin \phi + \epsilon \cos \phi + A(v/H)^{2.04}] G^{50}}{[c(\alpha+1) - \alpha v] E} dv \right\}, \quad (21)$$

$$s = \frac{400,000}{\sin \phi} \times \left[1 - \left(1 - \frac{v_0 t \sin \phi (k_1 \cos \phi)^{\frac{1}{2}}}{15,689} \right)^{1/25.5} \right], \quad (15)$$

$$\frac{dv}{dt} = \frac{dv}{ds} \frac{ds}{dt} = \frac{dv}{ds} v = \left[\frac{v_0^2 k_1 \sin 2\phi}{32,640} \right] \left[1 - \frac{s \sin \phi}{400,000} \right]^{-50}. \quad (16)$$

Substitute (16) into (11) and compare (11) with (3)

$$\left(\frac{1}{g}\right) k [c + \alpha(c-v)] = \sin \phi + \epsilon \cos \phi + \left[\frac{v_0^2 k_1 \sin 2\phi}{32,640g} \right] \left[1 - \frac{s \sin \phi}{400,000} \right]^{-50}. \quad (17)$$

Substitute (15) into (17) and obtain

$$k = \left[\frac{g}{c(\alpha+1) - \alpha v} \right] [\sin \phi + \epsilon \cos \phi + AB^{-1.96}], \quad (18)$$

$$A = \frac{v_0^2 k_1 \sin 2\phi}{32,640g} = \frac{E}{g}, \quad (18a)$$

$$B = 1 - \frac{v_0 t \sin \phi (k_1 \cos \phi)^{\frac{1}{2}}}{15,689}. \quad (18b)$$

Substitute (13) into (14) and obtain

$$t = \left[\frac{15,689}{v_0 \sin \phi (k_1 \cos \phi)^{\frac{1}{2}}} \right] \times \left[1 - \left(\frac{v_0 (k_1 \cos \phi)^{\frac{1}{2}}}{v} \right)^{1.04} \right]. \quad (19)$$

By the use of (5), the final expression for the mass ratio at the end of climbing is

or

$$i = \left(1 - \frac{s_1 \sin \phi}{400,000}\right), \quad (21a)$$

where v_1 is the velocity at the end of climbing:

$$v_1 = HG^{-24.5}, \quad (22)$$

$$l_1 = s_1 \cos \phi = h \cot \phi, \quad (23)$$

where the symbols denote

l_1 , horizontal distance traversed during climbing.
 s_1 , total length of flight path during climbing.

2. HORIZONTAL FLIGHT

To obtain the equations of horizontal flight, put in the above formulas the following values:

$$h = \text{const.}, \quad \phi = 0, \quad v = v_h.$$

As the initial values for this flight assume

$$W = W_1, \quad s = s_1, \quad v = v_1';$$

and as the end values:

$$t = t_2, \quad W = W_2, \quad s = s_2, \quad v = v_2.$$

Assume $\epsilon = \text{const.}$

$$\frac{W}{W_1} = \left(\frac{v}{v_1'}\right)^2, \quad \frac{W}{W_1} = \frac{W_1 + W_2}{2W_1} = k_2 = \text{const.}, \quad (24)$$

$$W_1 = \frac{\rho_h}{2} c_L S v_1'^2,$$

$$v = \frac{ds}{dt} = v_1'(k_2)^{\frac{1}{2}}, \quad (t - t_1) = \left[\frac{s - s_1}{v_1'(k_2)^{\frac{1}{2}}} \right], \quad (25)$$

$$(s - s_1) = v_1'(k_2)^{\frac{1}{2}}(t - t_1), \quad \frac{dv}{dt} = 0, \quad (26)$$

or

$$k = \left[\frac{g\epsilon}{c(\alpha + 1) - \alpha v} \right], \quad (27)$$

$$\begin{aligned} \frac{W}{W_1} &= \exp \left(- \int_{t_1}^t k dt \right) = \exp \left[- \frac{g\epsilon(t - t_1)}{c(\alpha + 1) - \alpha v} \right] \\ &= \exp \left[- \frac{g\epsilon(t - t_1)}{c(\alpha + 1) - \alpha v_1'(k_2)^{\frac{1}{2}}} \right]. \end{aligned} \quad (28)$$

At the end of horizontal flight one has

$$\begin{aligned} \frac{W_2}{W_1} &= 2k_2 - 1 = \exp \left[- \frac{g\epsilon(t_2 - t_1)}{c(\alpha + 1) - \alpha v_1'(k_2)^{\frac{1}{2}}} \right] \\ &= \exp \left[- \frac{g\epsilon(s_2 - s_1)}{v_1'(k_2)^{\frac{1}{2}} c(\alpha + 1) - \alpha v_1'^2 k_2} \right], \quad (29) \\ (s_2 - s_1) &= v_1'(k_2)^{\frac{1}{2}}(t_2 - t_1). \end{aligned} \quad (30)$$

3. GLIDING

During gliding there is no reactional force, and the weight of the projectile remains unchanged. In the case where the flight path is a curved line and the velocity of flight is very great, one has to take into account the centrifugal force. In flight with supersonic velocity the lift coefficient is a complicated function of several parameters, but as Sanger showed it may be given in a first very rough approximation by the formula

$$c_{LS} = \left(\frac{165,300}{v^2} \right) + 0.01. \quad (31)$$

Assume as the initial values $t = t_2$, $v = v_2$, $s = s_2$; and as the end values $t = t_3$, $v = v_3$, $s = s_3$. Apply the following symbols:

$$F_c = Wv^2/(Rg), \quad (32)$$

$$L = \frac{\rho_0}{2} S \left(\frac{165,300}{v^2} + 0.01 \right) \left(1 - \frac{h}{400,000} \right)^{49} v^2, \quad (33)$$

$$D = L\epsilon, \quad W = \text{const.}, \quad T = (W/g)(dv/dt), \quad (34)$$

where

F_c , centrifugal force;
 T , inertia force in the tangential direction;
 R , average radius of the earth.

Use the substitution:

$$p = \rho_0 S / (2W) = \text{const.}, \quad \epsilon = \text{const.}, \quad (35)$$

$$L = pW \left(\frac{165,300}{v^2} + 0.01 \right) \left(1 - \frac{h}{400,000} \right)^{49} v^2. \quad (36)$$

The equilibrium of forces in the radial and tangential directions gives two conditions:

$$\begin{aligned} p \left(\frac{165,300}{v^2} + 0.01 \right) \left(1 - \frac{h}{400,000} \right)^{49} v^2 \\ + \left(\frac{v^2}{Rg} \right) = \cos \phi, \end{aligned} \quad (37)$$

$$\epsilon \phi \left(\frac{165,300}{v^2} + 0.01 \right) \left(1 - \frac{h}{400,000} \right)^{40} v^2 = \sin \phi + \frac{1}{g} \left(\frac{dv}{dt} \right), \quad (38)$$

where ϕ is the instantaneous angle between the tangent to the path at an arbitrary point and the horizontal. It is assumed that the motion is decelerated.

Eliminating h from these two equations

$$\frac{d^2 s}{dt^2} = \frac{dv}{dt} = g \epsilon \cos \phi - g \sin \phi - \frac{v^2}{R}. \quad (39)$$

In order to solve this equation in an elementary way assume for ϕ a constant mean value. The difference between the straight flight path and the curved one is that the term representing the centrifugal force remains. After the first integration one obtains

$$t - t_2 = \left(\frac{1}{2K} \right) \ln \left[N \left(\frac{M + vK}{M - vK} \right) \right], \quad (40)$$

$$K = \left(\frac{\epsilon^2}{R} g \cos \phi - \frac{\epsilon}{R} g \sin \phi \right)^{\frac{1}{2}}, \quad (40a)$$

$$N = (M - v_2 K) / (M + v_2 K), \quad (40b)$$

$$M = g \epsilon \cos \phi - g \sin \phi; \quad (40c)$$

or

$$v = \left(\frac{M}{K} \right) \frac{\{\exp [2(t - t_2)K] - N\}}{\{\exp [2(t - t_2)K] + N\}}. \quad (41)$$

The second integration gives

$$(s - s_2) = \left(\frac{1}{K^2} \right) \ln \left\{ \frac{N + \exp \left[2 \frac{t - t_2}{K} \right]}{N + 1} \right\} - \left(\frac{M}{K} \right) (t - t_2). \quad (42)$$

The horizontal distance covered during this period may be calculated as a first approximation from the value $(s - s_2)$ and from the chosen value of the angle ϕ .

If the flight path is assumed to be a straight line one obtains from (39)

$$v - v_2 = g(\epsilon \cos \phi - \sin \phi)(t - t_2), \quad (43)$$

$$s - s_2 = (g/2)(\epsilon \cos \phi - \sin \phi)(t - t_2)^2 + v_2(t - t_2). \quad (44)$$

The horizontal distance during gliding is

$$l_2 = (s - s_2) \cos \phi. \quad (45)$$

In case the velocity of flight is smaller than the velocity of sound the coefficient ϵ_L is constant. Another way to approximately calculate the horizontal distance is

$$l_2 = K_2/D = K_2/W_2, \quad (46)$$

where K_2 is the kinetic energy at the end of horizontal flight. Assuming for D or ϵ a mean value during gliding one may obtain by short calculation the desired result.

Example

For the sake of simplicity only the distance covered in the horizontal flight period will be calculated. Assume the coefficient of internal efficiency of the motor to be 0.25, which is fairly high. As is known this coefficient is given by the expression

$$\eta_i = c^2/c_t^2, \quad (47)$$

where

c , real velocity of the exhaust gases.

c_t , theoretical velocity of the exhaust gases.

Consequently in the formulas employed, the real velocity c i.e., $c = 0.5c_t$, will be included instead of the theoretical velocity.

Use formula (29) with

$$\alpha = 15, \quad \epsilon = \frac{1}{5}, \quad c_t = 10,000 \text{ ft./sec.}, \\ v_1' = 300 \text{ ft./sec.}$$

* The tests of several airfoils at high speeds showed the following limits of ϵ at $V/c_t = 1.08$ (c_t is the speed of sound at temperature of jet) and at 0° or 2° of angle of attack relatively: RAF Family, 0.282 to 6.2 and 0.185 to 3.030, Clark Y Family, 0.318 to 4.530 and 0.167 to 3.290, Reed airfoil 0.698 and 0.435, Flat Plate 3.08 and 0.42. (L. J. Briggs and H. L. Dryden, "Aerodynamic characteristics of twenty-four airfoils at high speeds," N. A. C. A. Tech. Report No. 319, Washington, D. C.) The maximum theoretical velocity of the exhaust gases is $c_t = 9610(Q)^{\frac{1}{2}}$ ft./sec., where Q denotes the number of kilocalories per gram of mixture. The value of Q for several substances is: acetaldehyde 1.32, benzene 2.46, methane 2.63, normal hydrogen and normal oxygen 3.21, normal hydrogen and monoatomic oxygen 7.73, monoatomic hydrogen and monoatomic oxygen 13.35. (Aldo Vieira Da Rosa, "Notes of rocket fuels," J. Am. Rocket Soc., No. 61, 4-6 (March, 1945).) The hydrogen oxygen mixtures with Q greater than 3.21 are very unstable and they have not, as yet, been used.

The values of α and c_t may refer to gasoline. The result is

W_2/W_1	1:50	1:100	1:200	1:300
(s_2-s_1) in miles	1897	2220	2558	2745

As is seen, long distance bombing or mailing demands a high mass ratio. For $c = 30,000$ ft./sec., $\alpha = 15$, $\epsilon = \frac{1}{2}$, one obtains

W_2/W_1	1:50	1:100	1:200	1:300
(s_2-s_1) in miles	25,900	30,380	34,800	37,420

A final remark should be made concerning

formulas (7) and (27) given here in the original form in which they appeared in the literature. These formulas were derived for h in meters and v in m/sec. For calculations in ft. and ft./sec. one has to change the denominator of the second term inside the bracket in (7), i.e., to multiply it by 3.28083 and the numerator of the first term in (27), i.e., to multiply it by 10.763845. All the calculations based on these formulas should be adjusted to the new dimensions, i.e., the numerical coefficients in those formulas will be subject to a change.

Design of Broad Band I.F. Amplifiers

RICHARD F. BAUM

Raytheon Manufacturing Company, Waltham, Massachusetts

(Received February 19, 1946)

This paper deals with the design of broad band I.F. amplifiers with single tuned circuits in t stages. It is found that the figures of merit (Q_n) of the individual circuits should be related to the Q of a reference circuit according to:

$$Q/Q_n = \sin [(2m+1)\pi/2t] \cdot \frac{\omega_0}{\omega_{0n}}, \quad m=0, 1, 2, \dots, (t-1).$$

Then by proper tuning either an oscillatory or a monotonic response may be obtained. The relative band width BW/f_0 and the gain tolerance d_0 within the band determines the value of Q . The minimum number of stages for a given minimum attenuation in the cut-off region depends only on the gain tolerance and on the desired kind of response. Gain maxima (attenuation minima) appear at frequency deviations $\Delta f_n^{\min}$ from middle band frequency f_0 given by:

$$2\Delta f_n^{\min}/BW = \cos [(2m+1)\pi/2t].$$

Their location depends only on the number of stages. The resonance deviations Δf_{0n} of the tuned circuits are proportional to $\Delta f_n^{\min}$ with a proportionality factor F dependent on t and d_0 . The circuit impedances are calculated from a prescribed gain or from the maximum attainable gain. A formula for the maximum gain band width product is derived. The result is extended to double tuned stages. A practical example is given.

(A) INTRODUCTION

OF several types of broad band I.F. amplifiers the types designated as "stagger tuned" will be considered here. They comprise a number of single tuned stages with distinct resonance frequencies and thus differ from the normal I.F. amplifier group using identical stages. The design problem is mainly the proper distribution of the resonance frequencies within the frequency band and the attribution of proper selectivities to the individual circuits.

Amplifiers of this type have been described by several authors and their results given for a restricted number of circuits in tabulated form. It is obvious that a design procedure derived from general formulas is preferable, provided they do not include too complicated functions which obscure the mutual dependence of the parameters. One advantage of the proposed design consists in the simplicity of the resulting formulas, which involve only trigonometric and hyperbolic functions of real variables.

The success of the method is based on the very common assumption that the detuning ϵ of a resonant circuit is proportional to the frequency deviation from the resonance frequency f_0 :

$$\epsilon = 2(f - f_0)/f_0,$$

whereas the exact expression reads:

$$\epsilon = f/f_0 - f_0/f.$$

This approximation restricts the correctness of all derived formulas to a limited range, because for a band width of about one-half the center frequency the error in the gain of the stages resonating at one end of the band reaches 10 percent at the other end. A slight detuning of the outside stages is needed for compensation. With a band width of one-quarter the mid-band frequency this error could be compensated. In the case of wider bands the method still furnishes a first approach to the correct design values.

With this approximation the response curve of the amplifier becomes a function expressed by a polynomial of even powers of ϵ . On the other hand, the most desirable response curve may be developed in a polynomial of the same power of ϵ . By equalizing the two polynomials all circuit parameters may be calculated. This method of attack was outlined by Feldtkeller¹ who used the method of equalizing the coefficients of equal powers of ϵ . His method leads to a system of non-linear equations, the solution of which seems impossible for more than four circuits. Furthermore, it does not lead to general design formulas.

If the validity of the polynomials is continued into the region of complex frequencies then they can be equalized by equalizing their complex roots. This method has been applied in the present case. Use is made of equivalent representation of the response curve derived from the Tschebyscheff function. The roots of this expression can be calculated easily and lead to a very comprehensive result.

Considered as given design quantities are the ratio of band width to mid-band frequency, the gain, the gain tolerance within the band, and a certain rate of cut-off. Alternatively the design can be based on maximum gain.

(B) POSSIBLE ATTENUATION CURVES

The following symbols are used:

- g_n = mutual conductance,
- R_n = shunt resistance,
- Q_n = figure of merit,
- ϵ_n = twice the relative detuning,
- $\omega = 2\pi$ times the frequency f ,
- $\omega_0 = 2\pi$ times the mid band frequency f_0 ,
- $\omega_{0n} = 2\pi$ times the resonant frequency f_{0n} ,
- s_n = normalized detuning,
- Δs_{0n} = normalized resonance displacement,
- s_L = normalized half-band width,
- l = number of tuned circuits,
- BW = band width
- Δf_0 = half the band width
- d = attenuation (Nepers),
- d_0 = gain tolerance within the band (Nepers),
- d_m = minimum required attenuation outside the band (Nepers),
- l = loss in amplification,
- $m = 0, 1, 2, \dots (l-1)$,
- D_p = possible response curve,
- D_r = required response curve, and
- G = gain.

The index n refers to the n th circuit of the amplifier. Without this index the quantities apply to a reference circuit.

The gain of a single tuned stage is

$$G_n = g_n R_n / (1 + jQ_n \epsilon_n), \quad (1)$$

$$|G_n| = g_n R_n \cdot (1 + Q_n^2 \epsilon_n^2)^{-1/2}, \quad (2)^2$$

$$= g_n R_n \cdot (1 + s_n^2)^{-1/2}. \quad (3)$$

Each tuned circuit is adequately defined by its shunt resistance, its Q_n and resonant frequency f_{0n} . If a number of circuits are considered, we may calculate these quantities for a reference circuit, which does not form part of the amplifier itself, and the ratio of the Q 's and the resonance frequency displacement between the circuit in question and the reference circuit. Then the expression for the gain becomes

$$|G_n| = g_n R_n [1 + (a_n' s + b_n)^2]^{-1/2}. \quad (4)$$

The additional coefficient a_n' defines the Q ratio and the coefficient b_n relates the resonant frequencies. The following relations exist:

¹ Dr. R. Feldtkeller, *Theorie der Rundfunkschaltungen* (S. Hirzel, Leipzig).

² Compare: W. W. Hansen, J. App. Phys. 16, 529 (Sept. 1945); H. Wallman, "Stagger tuned I.F. amplifiers," M.I.T. Report 524 (Feb. 1944).

$$Q_n = R_n / \omega_{0n} L_n, \quad Q = R / \omega_0 L; \quad (5)$$

$$\epsilon_n = 2(\omega - \omega_{0n}) / \omega_{0n}, \quad \epsilon = 2(\omega - \omega_0) / \omega_0; \quad (6)$$

$$s_n = Q_n \epsilon_n, \quad s = Q \epsilon; \quad (7)$$

$$a_n = Q_n / Q, \quad a_n' = a_n \omega_0 / \omega_{0n}; \quad (8)$$

$$b_n = a_n \Delta s_n; \quad (9)$$

$$\Delta s_n = Q 2(\omega_0 - \omega_{0n}) / \omega_{0n}. \quad (10)$$

All the detunings and frequency displacements appear in normalized form that is multiplied by Q . The reference circuit is assumed to resonate at the mid-band frequency f_0 .

In accordance with Fig. 1, the index $n=1$ and t is reserved to the stages resonating at the edges of the band, the index 2 and $(t-1)$ to the stages next to it, and so on. In the case of an uneven number of stages, the index $(t+1)/2$ will apply to the stage resonating at mid-band frequency.

The gain of t stages is

$$|G| = \prod_{n=1}^{n=t} g_n R_n [1 + (a_n' s + b_n)^2]^{-1/2}. \quad (11)$$

In the middle of the band s equals zero and the gain becomes

$$|G_0| = \prod_{n=1}^{n=t} g_n R_n [1 + b_n^2]^{-1/2}. \quad (12)$$

If all stages would tune at the same frequency, the gain would be

$$|G_0^0| = \prod_{n=1}^{n=t} g_n R_n.$$

The root factor thus defines a loss l_0 in gain because of the displacement of resonant frequencies

$$l^2 = \prod_{n=1}^{n=t} (1 + b_n^2) \quad (13)$$

which makes

$$|G_0| = (1/l) \prod_{n=1}^{n=t} g_n R_n, \quad (14)$$

$$|G| = l |G_0| \prod_{n=1}^{n=t} [1 + (a_n' s + b_n)^2]^{-1/2}. \quad (15)$$

The factor in parenthesis defines the attenuation

$$d = .5 \lg \prod_{n=1}^{n=t} [1 + (a_n' s + b_n)^2], \quad (16)$$

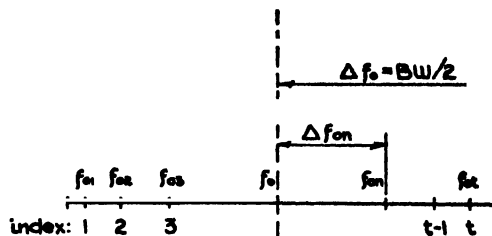


FIG. 1. Disposition of indexes and resonant frequencies. The lowest index $n=1$ is given the stage resonating at the lowest frequency f_{01} . The indexes increase with increasing resonance frequencies.

from which all possible response curves D_P follow with

$$D_P = \epsilon^{2d} = \prod_{n=1}^{n=t} [1 + (a_n' s + b_n)^2]. \quad (17)$$

D_P refers to the power response, whereas d refers to voltage attenuation and is measured in Nepers.

For reasons of symmetry with respect to the mid-frequency f_0 we anticipate the circuits to be symmetrically disposed with respect to f_0 , or

$$a_n' = a_{t-n+1}', \quad b_n = -b_{t-n+1}. \quad (18)$$

In this case the attenuation curve D_P is a polynomial in s containing only even powers of s and of a highest power $k_m = 2t$. This polynomial thus represents the general expression for all possible symmetrical response curves.

(C) REQUIRED RESPONSE CURVE

a. The Oscillatory Form

The required response curve has either an oscillatory form as in Fig. 2a and 2b, or a monotonic form as in Fig. 2c.

The former is defined by an oscillation of the gain between two fixed values. The magnitude of this oscillation shall be called "gain tolerance" d_0 and expressed in Nepers. The latter form shows a steady increase in attenuation without any oscillation. The required response curves are designated D_r .

Each minimum of the oscillatory response curve corresponds to one tuned circuit. Thus it shows t minima and $t-1$ maxima and can be represented by a polynomial of a highest power $k_m = 2t$ of the independent variable. Only even powers appear for reason of symmetry.

The band width shall be defined in a somewhat unusual way by that frequency deviation at

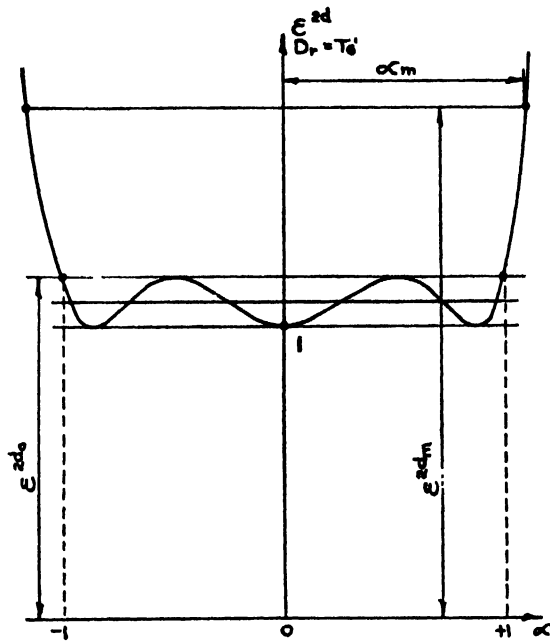


FIG. 2a

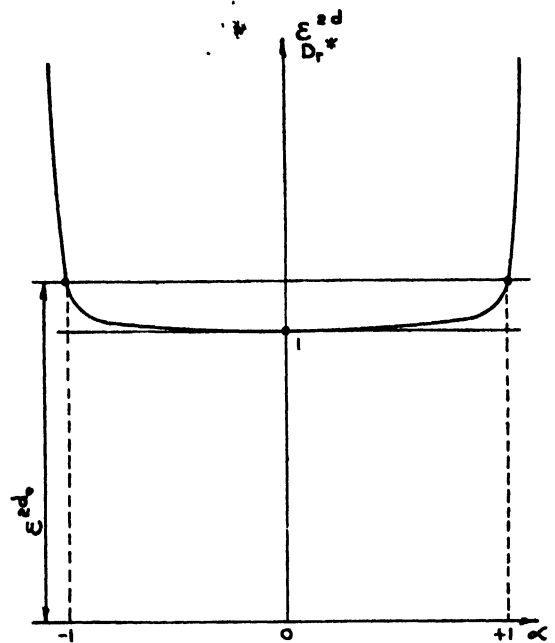


FIG. 2c

FIG. 2. Required response curves *a* and *b* show oscillatory, *c* monotonic behavior. The band width is reached for a prescribed attenuation d_0 . The tolerance d_0 is positive for an uneven number of stages (three in Fig. 2a) and negative for an even number (four in Fig. 2b).

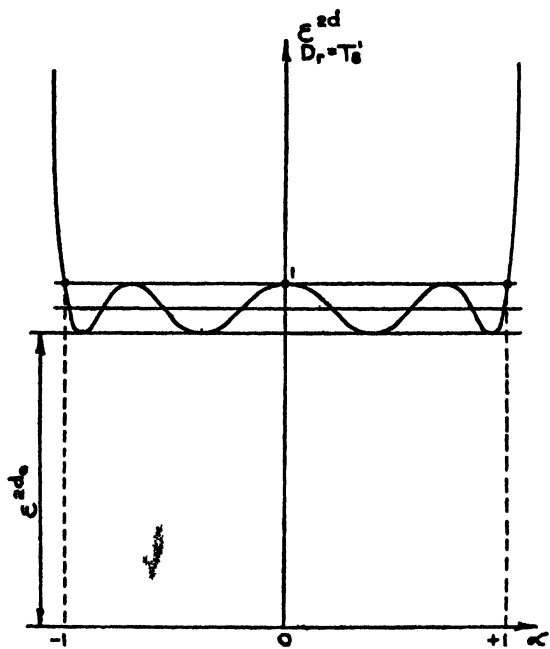


FIG. 2b

which the gain falls off a given amount of d_0 Nepers. In the oscillatory case d_0 coincides with the admissible gain tolerance. This band width will usually be narrower than the one defined by the half-power points.

With a normalized detuning s as abscissa

$$s = Q\Delta f/f_0$$

the half-band width appears in normalized form as

$$s_L = QBW/f_0. \quad (19)$$

In Fig. 2 the ratio

$$\alpha = s/s_L = 2\Delta f/BW \quad (20)$$

is taken as the independent variable; the band width then is reached for $\alpha = \pm 1$. At mid-frequency $\alpha = 0$ and the response equals one.

It is interesting to note that the polynomial of α of the $(2l)$ th order is completely defined by the demanded oscillatory behavior within the band $-1 < \alpha < +1$. Thus there is one definite position of the extremes within the band.

As proof, the roots of the derivative equation $dD_r/d\alpha$ may be extracted. They are functions of $(2l)$ coefficients. If introduced into the equation for D_r and if D_r is equalized to the prescribed extreme values, a system of $(2l-1)$ equations is found for $(2l+1)$ unknown coefficients. The remaining two equations derive from a prescribed value D_{r0} at the ends of the band.

If we can find another polynomial T_k' of α which within the same band shows the same oscillatory behavior as the polynomial D_r , the two have to be identical. The reason for introducing a new function T_k' will be apparent, when in the course of calculation the roots of $D_r=0$ have to be known. The roots of an equation of higher degree than the fourth may be expressed in general form only in special cases. With the aid of the new function T_k' it can be shown that the polynomial D_r belongs to such a special case. Furthermore T_k' allows writing down the coefficients of D_r almost without any calculation. Now these coefficients shall be given only as matter of interest; they are not useful for the present purpose.

The function T_k' is derived from the Tschebyscheff function T_k .³

$$T_k = \cos [k \cos^{-1} \alpha], \quad (21)$$

in which k is a parameter. Figure 3 shows this function for $k=6$. T_k oscillates between $+1$ and -1 within a band of $-1 < \alpha < +1$. It equals one for $\alpha = \pm 1$. The number of minima equals $k/2$. For values of $|\alpha| > 1$ it grows steadily and an equivalent expression in this region is

$$T_k = \cosh [k \cosh^{-1} \alpha]. \quad (22)$$

Within the band, T_k can be best described as a projection onto the $\alpha-T_k$ plane of a cos curve inscribed on the surface of a circular cylinder of unit radius with the T_k axis as its axis, k being the number of periods around the circumference. This explains the location of maxima, minima, and zeros according to Fig. 3 where the circumference of the unit circle is divided into $2k=12$ parts and the points thus found projected onto the α axis. The minima of T_k thus occur at

$$\alpha^{\min} = \cos [(2m+1)\pi/k]. \quad (23)$$

In order to show that T_k is nothing but a polynomial of α of a highest power α^k , it is best to use a recursion formula obtained as follows:

³ See for instance, E. Madelung, *Die mathematischen Hilfsmittel des Physikers* (Dover Publications, New York, 1943).

$$T_k = \cos (ky), \quad y = \cos^{-1} \alpha,$$

$$\cos [(k+1)y] + \cos [(k-1)y] = 2 \cos (ky) \cos y,$$

$$T_{k+1} + T_{k-1} = 2T_k T_1. \quad (24)$$

This formula allows one to compute T_{k+1} if T_k and T_{k-1} are known. Now as $T_0=1$ and $T_1=\alpha$ we obtain immediately

$$T_2 = 2\alpha^2 - 1, \quad T_3 = 4\alpha^3 - \alpha, \\ T_4 = 8\alpha^4 - 8\alpha^2 + 1, \quad \text{etc.}$$

The coefficients obey the law

$$A_{k,m} = 2A_{k-1,m-1} - A_{k-2,m}, \quad (25)$$

where $A_{k,m}$ is the coefficient of α^m in the polynomial T_k . This follows from (24).

The desired response curve D_r , Fig. 2a, differs from T_k only by a reduced magnitude of oscillation and the fact that D_r always equals $+1$ for $\alpha=0$, provided we make k equal to twice the number of stages. Thus an expression of the form

$$T_k' = pT_k + q, \quad (26) \\ 0 < p < 1,$$

$$k = 2t \quad (27)$$

can be made identical with D_r by proper choice of the two coefficients p and q . Neither the position of the extremes nor the power of the polynomial is affected.

The magnitude of oscillation is restricted by the maximum allowed gain variation within the

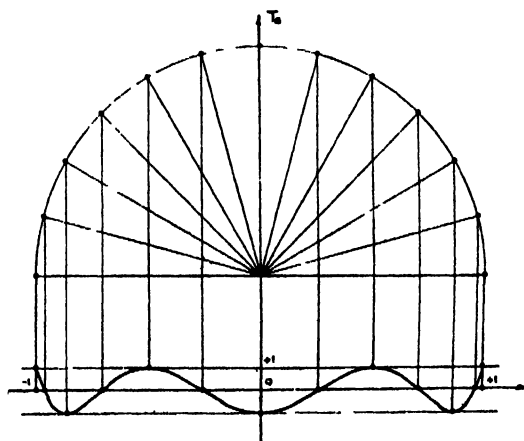


FIG. 3. The Tschebyscheff function $T_k(\alpha)$ for $k=6$. The position of maxima, minima, and zeros is evident from this construction. The number of minimas equals $k/2$. $T_k(\alpha)$ oscillates between $+1$ and -1 and reaches $+1$ for $\alpha = \pm 1$.

band. It shall be called gain tolerance d_0 (Nepers). Now there is a slight difference in T_k according to the value of $k=2t$, as for an even number of stages

$$T_k(0) = +1$$

and for an uneven number of stages

$$T_k(0) = -1.$$

The requirement then is

$$T_k'(0) = 1 \quad \text{for} \quad T_k(\alpha) = T_k(0)$$

and

$$T_k' = \exp(2d_0) \begin{cases} \text{for } T_k = +1 \text{ if it is even,} \\ \text{or } T_k = -1 \text{ if it is uneven.} \end{cases}$$

Furthermore, for an even number of stages the attenuation falls below zero level and d_0 has to be taken negative. The two cases are compared in Fig. 2a and b for $2t=6$ and $2t=8$, respectively. As we consider only small values of d_0 we may set

$$\exp(2d_0) = 1 + 2d_0 \quad (28)$$

obtaining

$$q = 1 + d_0, \quad p = |d_0|, \quad (29)$$

$$D_r = T_{2t}' = |d_0| T_{2t}(\alpha) + (1 + d_0). \quad (30)$$

As mentioned before, the purpose of writing the polynomial f or the required attenuation by means of the Tschebyscheff function was the possibility of expressing the roots of $D_r=0$ by a general formula. As these roots are complex, it has to be assumed that (27) is valid within the complex region if it is defined on the real α -axis. The proof of this shall not be given here.

$$D_r = |d_0| \cos[2t \cos^{-1} \alpha_0] + (1 + d_0) = 0, \\ 2t \cos^{-1} \alpha_0 = \cos^{-1} [-(1 + d_0)/d_0]. \quad (31)$$

The value of $(1 + d_0)/d_0$ is always larger than one. Let an auxiliary constant c be defined by

$$\cosh c = (1 + d_0)/d_0 \quad (32)$$

or

$$\cos(jc - \pi) = -(1 + d_0)/d_0.$$

Because of the periodicity of the cos function the real part of its argument is multiple valued with a period of 2π :

$$\cos(jc - \pi - 2m\pi) = -(1 + d_0)/d_0, \\ m = 0, 1, 2, \dots$$

This introduced into Eq. (31) gives for the roots of D_r :

$$\alpha_0 = \cos \left[j \frac{c}{2t} - \frac{2m+1}{2t} \pi \right] \quad (33)$$

and for their real and imaginary part

$$\alpha_R = \cosh(c/2t) \cos[(2m+1)\pi/2t], \quad (34)$$

$$\alpha_I = \sinh(c/2t) \sin[(2m+1)\pi/2t]. \quad (35)$$

b. The Monotonic Form

If in the polynomial defined by Eq. (17) all coefficients but one are made equal to zero, the polynomial reduces to an expression of the form

$$D_P^* = A + Bs^k.$$

This suggests the second form of response curve D_P as sketched in Fig. 2c, which shows a flat base within the band. In order to obtain the flattest response together with the steepest cut-off slope, k has to be chosen as large as possible, that is equal to $2t$.

The half-band width in this case shall be defined by the value of $s=s_L$ at which the attenuation attains a prescribed value d_0 (Nepers). Again $\alpha=s/s_L$ is used as the variable and zero attenuation is required at mid-band frequency ($\alpha=0$), which gives with the approximation of Eq. (28)

$$D_r^* = 1 + 2d_0\alpha^{2t}, \quad (36)$$

as the expression for the required monotonic response curve.

The roots of this expression can easily be found

$$\alpha_0^{*2t} = -1/2d_0 = \exp[j(\pi + 2m\pi)]/2d_0, \\ \alpha_0^* = [2d_0]^{-1/2t} \exp[j(2m+1)\pi/2t]; \quad (37)$$

their real and imaginary parts are

$$\alpha_R^* = [2d_0]^{-1/2t} \cos[(2m+1)\pi/2t], \quad (38)$$

$$\alpha_I^* = [2d_0]^{-1/2t} \sin[(2m+1)\pi/2t]. \quad (39)$$

The close relationship of these expressions to Eqs. (34) and (35) should be noted. They contain the same cos and sin factors. The cos factor again coincides with the expression for the location of

⁴All quantities referring exclusively to monotonic response are distinguished by a star. The solution of this case is given in Wallman's paper, reference 2.

the attenuation minima, according to Eq. (23).

$$\alpha^{\min} = \cos [(2m+1)\pi/2t]. \quad (40)$$

The physical meaning of these resemblances shall be discussed in the next paragraph.

(D) CALCULATION OF CIRCUIT PARAMETERS

a. Selectivities and Resonant Frequencies

By equalizing the roots given by Eq. (37) or Eq. (33) with the roots of Eq. (17), all necessary relationships between circuit parameters are obtained. The roots of Eq. (17) are

$$s_{0n} = b_n/a_n' + j/a_n'; \quad (41)$$

by division with s_L :

$$\alpha_{0n} = b_n/a_n' s_L + j/a_n' s_L. \quad (42)$$

By comparing Eq. (42) and Eq. (33) we obtain for the oscillatory case

$$1/a_n' = s_L \sinh(c/2t) \sin[(2m+1)\pi/2t], \quad (43)$$

$$b_n/a_n' = s_L \cosh(c/2t) \cos[(2m+1)\pi/2t]. \quad (44)$$

The value of s_L now can be fixed by demanding that for an uneven number of circuits the reference circuit should be identical with the circuit resonating at mid-band frequency, that is by requiring $a_n' = 1$, $b_n = 0$ for $m = (t-1)/2$. This is possible only if:

$$s_L = 1/\sinh(c/2t). \quad (45)$$

Together with Eqs. (8) and (9)

$$1/a_n' = (Q/Q_n)(\omega_{0n}/\omega_0) = \sin[(2m+1)\pi/2t], \quad (46)$$

$$b_n/a_n' s_L = (\Delta f_{0n}/\Delta f_0) = \alpha^{\min} \cosh(c/2t), \quad (47)$$

where Δf_{0n} is the displacement of the resonance frequency from the mid-band frequency and Δf_0 is half the band width, as shown in Fig. 1.

Between m and n a relation exists of the form

$$m = n - 1, \quad (48)$$

which means that by setting successively m equal to $m = 0, 1, 2, 3, \dots, (t-1)$ all circuit parameters for $n = 1, 2, 3, \dots, t$ are obtained.

The same reasoning in the case of a monotonic response curve leads to the formula

$$s_L^* = [2d_0]^{1/2t}, \quad (49)$$

$$(Q^*/Q_n^*)(\omega_{0n}^*/\omega_0) = \sin[(2m+1)\pi/2t], \quad (50)$$

$$\Delta f_{0n}^*/\Delta f_0 = \alpha^{\min} \cdot [2d_0]^{-1/2t}. \quad (51)$$

Equation (50) is identical with Eq. (46). For the following discussion, Eq. (40) is joined in the form

$$\alpha^{\min} = \frac{\Delta f_n^{\min}}{\Delta f_0} = \cos[(2m+1)\pi/2t]. \quad (52)$$

It should be noted that, in the first approximation, the factor ω_{0n}/ω_0 and ω_{0n}^*/ω_0 equals one; but that in the case of wide bands and especially for the circuits tuning at the cut-off region, this factor may be considerably different from 1. The asymmetry it introduces shows an increase in Q when passing through the band from the low to the high frequency end.

The best insight into the result is obtained by using the first approximation only and may be stated now this way:

In order to achieve either a monotonic or an oscillatory type of response for a given number of t circuits, the selectivities of the individual circuits have to be staggered according to the law

$$Q/Q_n \approx \sin[(2m+1)\pi/2t]. \quad (53)$$

Thus the circuit tuning nearest to the mid-band frequency f_0 receives the lowest Q and the further a circuit tunes off f_0 the higher a Q it obtains.

The circuits of a monotonic amplifier should resonate according to Eq. (51) at frequency displacements proportional to the displacements, where an amplifier with oscillatory response has its gain peaks, that is at resonance displacements

$$\Delta f_{0n}/\Delta f_0 = [2d_0]^{-1/2t} \cos[(2m+1)\pi/2t]. \quad (54)$$

From this expression it is apparent that the resonance points crowd more and more together as one moves from the middle to the edges of the band. This behavior as well as the increased selectivity could have been predicted from the general appearance of the Tschebyscheff function.

For an amplifier of oscillatory response the resonance displacement is larger by a proportionality factor F

$$F = \cosh(c/2t), \quad (55)$$

where

$$\cosh c = (1+d_0)/|d_0|.$$

The larger the admissible gain variation d_0 and the larger the number of stages, the better the resonance frequencies and the peak frequencies will coincide.

The absolute value of selectivity follows from Eq. (19) with

$$Q = s_L f_0 / BW.$$

For a given relative band width BW/f_0 thus

$$QBW/f_0 = s_L. \quad (56)$$

The larger the relative band width the lower the Q of the circuits has to be. For a given relative band width for oscillatory (Eq. (45)) or monotonic response (Eq. (49)) Q increases with the gain tolerance d_0 .

b. The Minimum Number of Circuits

The minimum number of circuits can be calculated if a minimum attenuation of d_m (Nepers) at a given frequency displacement from mid-band frequency is required, as shown in Fig. 2a. If this displacement equals α_m times the half-band width, then with Eq. (22) and (30)

$$|d_0| \cosh [2t \cosh^{-1} \alpha_m] + (1 + d_0) \geq \epsilon^{2d_m},$$

$$\cosh [2t \cosh^{-1} \alpha_m] \geq [\epsilon^{2d_m} - (1 + d_0)] / |d_0|.$$

Introducing an auxiliary quantity h by

$$\cosh h = [\epsilon^{2d_m} - (1 + d_0)] / |d_0|, \quad (57)$$

and solving for t

$$2t \geq h / \cosh^{-1} \alpha_m. \quad (58)$$

The minimum number of stages thus increases with the required attenuation d_m but decreases with the admissible gain variation d_0 . In most cases $\epsilon^{2d_m} \gg (1 + d_0)$ and Eq. (57) can be reduced to

$$\cosh h \cong \epsilon^{2d_m} / |d_0|, \quad (59)$$

and further to

$$h \cong 2d_m - \lg(d_0/2). \quad (60)$$

On the other hand a given attenuation will be reached for frequencies given by

$$\alpha_m' = \cosh(h/2t). \quad (61)$$

The corresponding formulas for the monotonic case, using Eq. (36), are $1 + 2d_0\alpha_m^{2t} \geq \epsilon^{2d_m}$

$$h^* = \lg[(\epsilon^{2d_m} - 1)/2d_0] \quad (62)$$

and

$$2t^* \geq h^* / \lg \alpha_m. \quad (63)$$

Equation (62) again most frequently reduces to

$$h^* \cong 2d_m - \lg(2d_0). \quad (64)$$

Equations (57) to (64) should be used with caution as they refer to the behavior of the response curves outside the band width. It was pointed out before that in this region the assumption for ϵ (according to Eq. (6)) becomes increasingly incorrect. It is well to restrict α_m to values $1 < \alpha_m < 2$. Also if the required minimum attenuation d_m is taken very large compared with the gain tolerance, that is if

$$2d_m \gg \lg d_0,$$

the minimum number of stages becomes proportional to d_m . Then it would not make much difference if the amplifier has t stages or if it is broken up, for instance into two identical parts of $t/2$ stages each. This is another reason to restrict α_m to small values, where the attenuation is not so large.

By means of (58) and (63), it can be shown that within the restricted region of α_m breaking up of the amplifier into two or more parts leads to a greater number of stages for the same required attenuation d_m and the same tolerance d_0 . This statement, in other words, says that a single unit amplifier designed according to the present scheme has a better cut-off characteristic than any other amplifier of the same number of stages either identical or identical in pairs or triplets or more sub-units. Indeed for the single unit amplifier (with (58) and (60))

$$2t = [2d_m - \lg(d_0/2)] / \cosh^{-1} \alpha_m.$$

If this amplifier is composed of i identical parts, then each has to have a minimum attenuation d_m/i and a tolerance d_0/i . The total number of stages necessary is:

$$i \cdot 2t_i = i[2d_m/i - \lg(d_0/2i)] / \cosh^{-1} \alpha_m.$$

The previous statement is correct if always

$$i[2d_m/i - \lg(d_0/2i)] > 2d_m - \lg(d_0/2)$$

or if

$$[d_0/2]^{i-1} < i;$$

this is correct, as always

$$i > 1 \text{ and } d_0 < 1.$$

The analogous proof in the case of monotonic response leads to the condition:

$$[2d_0]^{i-1} < i^i$$

which is also true.

This result represents a justification for the choice of the required response curves. Another one is derived from the property of the Tschebyscheff function $T_k(\alpha)$ of being the one polynomial of highest exponent k and smallest coefficient $+1$ or -1 which within the band $-1 < \alpha < +1$ approaches closest zero. Any other polynomial of the same order k attains a larger value at least at one point within the band. The chosen response curve D_r thus is the curve which within the band most closely approaches constant gain, if the number of circuits is prescribed.

c. Impedances and Gain

The staggering of resonance frequencies results in a loss l in gain at the mid-band frequency given by Eq. (13). By dividing Eq. (44) by Eq. (43), l becomes

$$l^2 = \prod_{m=0}^{m=i-1} [1 + \coth^2(c/2t)] \times \cot^2[(2m+1)\pi/2t] \quad (65)$$

in the oscillatory case, whereas in the monotonic case

$$l^{*2} = \prod_{m=0}^{m=i-1} [1 + \cot^2[(2m+1)\pi/2t]]$$

or

$$l^* = \prod_{m=0}^{m=i-1} [1/\sin[(2m+1)\pi/2t]] = 2^{i-1}. \quad (66)$$

Each additional stage here reduces the gain by a factor of 2.

In order to calculate the gain, an assumption about the shunt resistance R_n has to be made. The stages of the amplifier will most likely be designed either with the same capacity C and inductively tuned, or with the same inductance L and capacitively tuned. In the former case

$$\begin{aligned} Q_n &= C\omega_n R_n, & Q &= C\omega_0 R, \\ R_n/R &= Q_n\omega_0/Q\omega_n = a_n'. \end{aligned} \quad (67)$$

In the latter case

$$\begin{aligned} Q_n &= R_n/\omega_n L, & Q &= R/\omega_0 L, \\ R_n/R &= Q_n\omega_n/Q\omega_0 = a_n'(\omega_n/\omega_0)^2. \end{aligned} \quad (68)$$

Only the first case shall be considered here, as it gives a slightly larger gain. The corresponding equations for the other case can be derived easily. Introducing Eq. (67) into the gain formula (14)

$$|G_0| = (g^t R^t/l) \prod_{n=1}^{n=i} a_n' \quad (69)$$

and with

$$\begin{aligned} \prod_{n=1}^{n=i} a_n' &= \prod_{m=0}^{m=i-1} [1/\sin[(2m+1)\pi/2t]], \\ &= L^* = 2^{i-1}, \end{aligned} \quad (70)$$

$$|G_0| = (2gR)^t/2l. \quad (71)$$

Identical tubes have been assumed. The first stage often consists of an input transformer with a step-up ratio g_r . Then g^t would have to be replaced by $g_r g^{t-1}$.

In some cases by reason of the large number of tubes involved the required gain is easily achieved and R may be calculated from Eq. (71)

$$\log_{10}(2gR) = [\log_{10}|G_0| + \log_{10} 2l]/t. \quad (72)$$

For the monotonic response the gain expression becomes extremely simple because of the relation (70); namely,

$$|G_0^*| = (gR)^t. \quad (73)$$

The gain of the monotonic amplifier is the same as the gain of an amplifier consisting only of reference circuits, all tuned at mid-frequency.

Gain formula (71) involves the loss l , which has to be calculated by means of Eq. (65). It is a function of $c/2t = \gamma$. In order to facilitate the practical design $\log_{10}(2l)$ is drawn in Fig. 4 as a function of t and γ .

The gain can as well be represented as a function of the capacity C . With

$$Q = RC\omega_0$$

and

$$QBW/f_0 = s_L$$

the shunt resistance equals

$$R = s_L f_0 / C\omega_0 BW \quad (74)$$

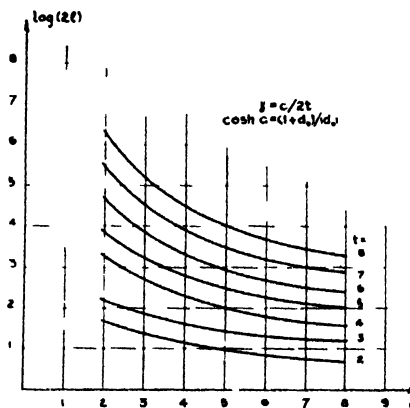


FIG. 4. Value of $\log(2l)$ as a function of γ and the number of stages t . The loss in gain l is caused by the staggering of the resonant frequencies.

and the gain

$$|G_0| = [(g/\pi C)(s_L/BW)]^t/2l \quad (75)$$

and

$$|G_0^*| = [(g/\pi C)(s_L^*/2BW)]^t, \text{ respectively.} \quad (76)$$

The maximum gain is obtained by setting C equal to the sum of input and output capacities of the tube

$$C = C_{in} + C_{out} = C_m.$$

The maximum average gain band width product follows with

$$|\bar{G}_0| BW = (g/\pi C_m) s_L (2l)^{-1/t} \quad (77)$$

and

$$|\bar{G}_0^*| BW = (g/\pi C_m) s_L^*/2. \quad (78)$$

In the quantity in parenthesis the theoretical maximum is recognized, obtainable by optimum distribution of an infinite number of circuits within the band.⁵ The additional factor then will give a clue for the estimation of the gain efficiency of the presented method. It should not be forgotten though that the theoretical maximum is based on constant gain within the band and zero tolerance for the definition of the band width, whereas for a limited number of circuits the band width has to be defined by a given finite tolerance.⁵

This additional factor μ

$$\mu = s_L (2l)^{-1/t}$$

⁵ See W. W. Hansen, reference 2.

is within the practical region for γ , namely, $.2 < \gamma < .8$, almost independent of the number of stages and given in Table 1, for 3 to 8 stages (with an accuracy of ± 5 percent). We thus find $.81 < \mu < .45$.

In the monotonic case

$$\mu^* = .5(2d_0)^{1/2t}$$

which is always smaller than one-half and approaches this value as the number of stages t increases.

(E) EXTENSION OF RESULTS

Instead of using only one single tuned circuit in each stage, we may use two or more tuned circuits either capacitively or magnetically coupled. With certain assumptions all of them lead to response curves represented by polynomials of ϵ . This equivalence is treated in a series of articles.⁶ The results obtained for stagger tuned single circuits apply just as well to these cases provided the calculated coefficients are properly interpreted.

Of special interest is the case of two magnetically coupled circuits. If the coupling is specified by a coupling coefficient n (equal to one for transitional coupling) the gain of one stage can be written

$$|G_n'| \cong [(1 + n^2 - s^2)^2 + 4s^2]^{-1/2} \cong [1 + (s + n)^2]^{-1/2} [1 + (s - n)^2]^{-1/2} \quad (79)$$

By comparing with the gain of two single tuned stages

$$|G_n''| = g^2 R_n^2 [1 + (a_n' s + b_n)^2]^{-1/2} [1 + (a_n' s - b_n)^2]^{-1/2},$$

it follows that the coefficient b_n now has to be interpreted as coupling coefficient. The distribution of coupling through the band thus is given by

$$n_n = \coth(c/2l) \cot[(2m+1)\pi/2l] \quad (80)$$

and

$$n_n^* = \cot[(2m+1)\pi/2l]. \quad (81)$$

Furthermore $a_n' = 1$ requires $RC = R_n C_n$. The number of stages now equals $2l$.

⁶ Maynard, Communications 38 (Jan. 1945); Gardiner and Maynard, Proc. I.R.E. 676 (Nov. 1944); R. Feldtkeller, reference 1.

(F) EXAMPLE

An example best shows the application of the derived formulas to a practical problem:

An intermediate frequency stagger tuned amplifier is to be designed having a mid-band frequency of 2 mc, a band width of .5 mc, and a gain of 100,000. The gain tolerance is $d_0 = .1$ Nepers and an attenuation of at least 20 db (or 2, 3 Nepers) has to be reached at 1, 3 times the band width. An oscillatory response is desired. Thus:

$$\begin{aligned} f_0 &= 2 \text{ mc}, \quad \Delta f_0 = .25 \text{ mc}, \quad |G_0| = 10^5, \\ d_0 &= .1, \quad d_m = 2.3, \quad \alpha_m = 1.3. \\ h &= 4.60 + .693 + 2.302 = 7.595, \quad (60) \\ 2l &\geq 7.595/.76 = 10; \quad (58) \end{aligned}$$

thus at least 5 stages are needed. Maximum gain occurs at:

$$\begin{aligned} \Delta f_n^{\min}/\Delta f_0 &= \pm .951 \text{ and } \pm .588 \text{ and } 0. \quad (52) \\ \Delta f_n^{\min} &= \pm .238 \text{ and } \pm .147 \text{ and } 0 \text{ mc.} \end{aligned}$$

The correction factors in (50) and (51) thus are approximately

$$f_{0n}/f_0 \cong f_n^{\min}/f_0 = .88, .926, 1, 1.073, \text{ and } 1.119.$$

The auxiliary parameters c and s_L give the selectivities:

$$c = \cosh^{-1} \frac{1.1}{.1} = \cosh 11 = 3.09, \quad (32)$$

$$\gamma = c/2l = .309, \quad s_L = 3.18, \quad (45)$$

$$Q = 3.18 \frac{2}{.5} = 12.72, \quad (56)$$

$$\frac{Q_{\omega_{0n}}}{Q_{\omega_0}} \cong \frac{Q_{f_n^{\min}}}{Q_{f_0}} = .309 \text{ and } .809 \text{ and } 1. \quad (46)$$

$$\begin{aligned} Q_1 &= 36.2, \quad Q_2 = 14, \quad Q_3 = 12.72 = Q, \\ Q_4 &= 46.1, \quad Q_5 = 16.3. \end{aligned}$$

The resonance displacements Δf_{0n} derive from

TABLE I. Value of μ for 2 to 8 stages for different values of γ . The maximum gain theoretically attainable corresponds to $\mu = 1$.

$\gamma = c/2l =$							
.2	.3	.4	.5	.6	.7	.8	
$\mu = .81$	.75	.675	.61	.55	.5	.45	

$$\cosh (c/2l) = \cosh .309 = 1.048,$$

$$\frac{\Delta f_{0n}}{\Delta f_0} = \pm .998 \text{ and } \pm .617 \text{ and } 0. \quad (47)$$

$$\left. \begin{aligned} \Delta f_{01} &= -.250, \quad \Delta f_{02} = -.154, \quad \Delta f_{03} = 0 \\ \Delta f_{04} &= +.250, \quad \Delta f_{05} = +.154. \end{aligned} \right\} \text{mc.}$$

The shunt resistances follow from the required gain:

$$\log_{10} (2gR) = [5 + 3.19]/5 = 1.64, \quad (72)$$

where $\log (2l)$ is taken from Fig. 4. With a transconductance $g = 5000 \mu\text{mhos}$,

$$R = 4370 \text{ ohms.}$$

$$R/R_n = 1/u_n' = .309 \text{ and } .809 \text{ and } 1. \quad (67) \text{ and } (46)$$

$$\left. \begin{aligned} R_1 &= R_5 = 10900 \\ R_2 &= R_4 = 5300 \\ R_3 &= R = 4370 \end{aligned} \right\} \text{ohms.}$$

Each circuit thus is specified by its shunt resistance, its Q and its resonance point.

(G) CONCLUSION

A method has been outlined for the design of broad band amplifiers, based on identification of the response curve with either the Tschebycheff polynomial or a parabola of higher degree. The former applies to an oscillatory and the latter to a monotonic response. The simplicity of the resulting formulas was achieved by an approximation which restricts their use to a limited range of band width. The method of attack lends itself just as well to the exact calculation (for any band width) and shall be published in another paper.

The Relation between Nodal Positions and Standing Wave Ratio in a Composite Transmission System*

EUGENE FEENBERG
New York University, New York, New York
(Received December 21, 1945)

Reflection generally occurs at a lossless transition region joining two uniform lossless lines. If the output line feeds into a matched load (no reflection) a standing wave ratio η_0 different from unity exists on the input side of the transition region. If the output line is terminated by a movable short circuit, a relation exists between the nodal positions on opposite sides of the transition section. The relation can be used to determine η_0 thus dispensing with the need for a calibrated detecting system to measure this quantity.

FOR the purpose of this note a composite transmission system is defined to be a transmission line composed of two uniform lossless sections (coaxial or wave guide) joined by any sort of lossless adapter. In particular the system may be simply a single uniform line containing a lossless irregularity (window, supporting bead, stub section, etc.) localized somewhere along the line. Systems of this general type are employed in many varieties of microwave equipment and measuring apparatus. Actual systems are not, of course, lossless, but it is generally true that power losses are kept to a minimum by good design.

In Fig. 1 the incident wave travels from left to right. Region 1 feeds into a matched load so that no reflection occurs from the output end of the system. However, reflection may occur at the adapter or irregularity resulting in nodes and a standing wave ratio η_0 different from unity in region 2.

If region 1 is terminated by a shorting plug the location of the nodes in both regions depends on the position of the shorting plug. We show that the standing wave ratio η_0 which exists in

region 2 when region 1 feeds into a matched load (no reflection) can be expressed simply in terms of the relation between the nodal positions in the two regions when region 1 is terminated by a shorting plug. This statement has the practical consequence that standing wave ratios can be evaluated without the use of a calibrated detecting system.

The analysis is based on the assumption that outside of a transition region centered about the origin the nodes are uniformly spaced. In each uniform section the distance between alternate nodes defines wave-lengths λ_1 in region 1 and λ_2 in region 2 which may or may not differ from each other and from the wave-length λ in free space. The assumption of uniformly spaced nodes does not exclude the possibility that the uniform sections are capable of transmitting more than one mode without attenuation, but does require that under the experimental conditions only one such mode is present in each uniform section.

Let E represent the voltage (in the case of a coaxial line) or the transverse electric field (in the case of a wave guide). Outside of the transition region the fields are given by the expressions

$$\begin{aligned} E_1 &= A \exp \{-jk_1(x-x_1)\}, \\ E_2 &= B[\exp \{-jk_2(x-x_2)\} \\ &\quad - b \exp \{jk_2(x-x_2)\}] \end{aligned} \quad (1)$$

for the case pictured in Fig. 1. Here A and B are real quantities which may depend on transverse coordinates, but are independent of x . The reflected amplitude b is real and positive. Then the voltage squared standing wave ratio in region 1 has the value

$$\eta_0 = (1+b)^2/(1-b)^2. \quad (2)$$

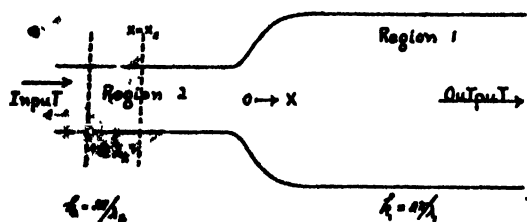


FIG. 1.

* Based on reports written for the Research Laboratory of the Sperry Gyroscope Company, Garden City, New York.

Figure 2 illustrates the field relations when region 1 is terminated by a shorting plug.

To construct solutions for this case we make use of the fact that the complex conjugate of the field defined by Eq. (1) is also a possible field

$$\begin{aligned} E_1' &= A[\exp\{-jk_1(x-x_1')\} - \exp\{jk_1(x-x_1')\}], \\ E_2' &= B \exp\{-jk_2(x-x_2)\}[\exp\{jk_1(x_1'-x_1)\} + b \exp\{-jk_1(x_1'-x_1)\}] \\ &\quad - B \exp\{jk_2(x-x_2)\}[\exp\{-jk_1(x_1'-x_1)\} + b \exp\{jk_1(x_1'-x_1)\}]. \end{aligned} \quad (3)$$

This solution vanishes at the short ($x=x_1'$) and thus represents the situation pictured in Fig. 2. In region 2 a nodal surface is located at the position

$$x_2' = x_2 + 1/k_2 \cdot \tan^{-1} [1/\eta_0 \cdot \tan k_1(x_1' - x_1)]. \quad (4)$$

Equation (4) can be transformed into

$$k_2(x_2' - x_2) = k_1(x_1' - x_1) - \tan^{-1} \left[\frac{(\eta_0^2 - 1) \tan k_1(x_1' - x_1)}{\eta_0^2 + \{\tan k_1(x_1' - x_1)\}^2} \right]. \quad (5)$$

The Eqs. (4) and (5) reduce to the linear relation

$$k_2(x_2' - x_2) = k_1(x_1' - x_1) \pm n\pi, \quad (6)$$

if the transition region produces no reflection ($\eta_0 = 1$).

A plot of the quantity $k_2x_2' - k_1x_1'$ is shown in Fig. 3. The range D over which $k_2x_2' - k_1x_1'$ varies is readily found to have the value

$$D = 2 \tan^{-1} \frac{1}{2} (\eta_0^2 - \eta_0^{-2}) \cong \eta_0^2 - \eta_0^{-2} \cong 2b \quad (\text{when } b \ll 1). \quad (7)$$

The maximum and minimum values of $k_2x_2' - k_1x_1'$ occur at values of x_1' given by the relation

$$\tan k_1(x_1' - x_1) = \pm \eta_0^2. \quad (8)$$

If η_0 is known, Eq. (7) supplies the answer to the question of how much the relation between the nodal displacements departs from linearity. Conversely an experimental determination of the relation between the nodal displacements yields a value for η_0 .

The phase k_1x_1 which appears in Eq. (1) is not readily accessible to measurement under the conditions of Fig. 1. However, a suitable linear combination of Eq. (1) and its conjugate complex yields a solution for the situation in which input

(the associated magnetic field or current quantity is replaced by the negative of its conjugate when E is replaced by E^*). Taking a suitable linear combination of the two fields at our disposal we obtain

and output are interchanged and no reflected wave occurs in region 2. This solution is

$$\begin{aligned} E_2'' &= KB(1-b^2)^{1/2} \exp\{jk_2(x-x_2)\} \\ E_1'' &= KA(1-b^2)^{-1/2} [\exp\{jk_1(x-x_1)\} \\ &\quad + b \exp\{-jk_1(x-x_1)\}], \end{aligned} \quad (9)$$

in which K is a constant. Thus nodes occur in region 1 at the positions

$$k_1x_1'' = k_1x_1 + (n_1 - \frac{1}{2})\pi, \quad n_1 = 0, \pm 1, \dots \quad (10)$$

when the source is located on the right (region 1) and region 2 feeds into a matched load (no reflection).

In the special case of a uniform line containing a reflecting element (which may occupy an arbitrary length of the line) Poynting's theorem yields the relation

$$A^2 = B^2(1-b^2). \quad (11)$$

With the help of Eq. (11) the double primed solution can be expressed in the form

$$\begin{aligned} E_2'' &= KA \exp\{jk_1(x-x_2)\}, \\ E_1'' &= KB[\exp\{jk_1(x-x_1)\} \\ &\quad + b \exp\{-jk_1(x-x_1)\}]. \end{aligned} \quad (12)$$

Now, assuming a *symmetrical* reflecting element, Eqs. (1) and (12) are related in the same way as object and image in a plane mirror. Applying the mirror image relation to the location of the

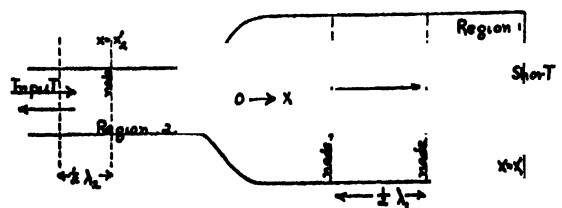


FIG. 2.

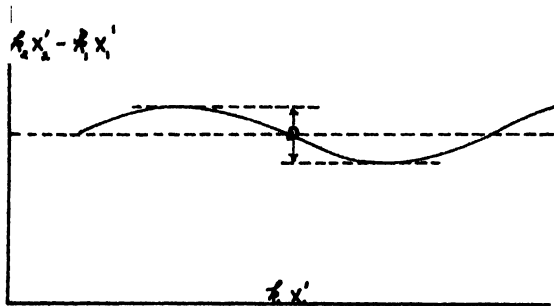


FIG. 3.

nodes we get

$$k_1 x_1'' = k_1 x_1 + (n_1 - \frac{1}{2})\pi \\ = -k_1 x_2 + n_2 \pi, \quad n_2 = 0, \pm 1, \dots \quad (13)$$

Consequently,

$$k_1(x_1 + x_2) = (n + \frac{1}{2})\pi, \quad n = 0, \pm 1, \dots \quad (14)$$

In the limiting situation produced by reducing

the irregularity to the vanishing point continuity requires

$$k_1 x_2 = k_1 x_1 + m\pi, \quad m = 0, \pm 1, \dots \quad (15)$$

or

$$k_1 x_2 = (2p + 1)\pi/4, \quad p = 0, \pm 1, \dots \quad (16)$$

Equation (16) can also be expressed in the form

$$x_2 = \lambda_1/8 + p\lambda_1/4. \quad (17)$$

There are evidently two physically distinct cases; one associated with even values of p , the other with odd values. One can show that an infinitesimal lumped capacitive element across the uniform line produces a node at $x_2 = -\lambda_1/8$ (when power flows from left to right). On the other hand a very large lumped inductive element across the line places a node at $x_2 = -3\lambda_1/8$. Thus the lumped capacitive case is associated with odd values of p and the lumped inductive case with even values.

Erratum: Small Prism Infra-Red Spectroscopy

[J. App Phys 16, 77-86 (1945)]

R. BOWLING BARNES, ROBERT S. McDONALD, AND VAN ZANDT WILLIAMS
Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut

AND

RICHARD F. KINNAIRD
The Perkin-Elmer Corporation, Glenbrook, Connecticut

E. GANZ, Basle, Switzerland, has called our attention to an error in our paper. The formula on page 81 should read

$$\Delta \nu (cm^{-1}) = \frac{A}{8 \sin \frac{\alpha}{2} \frac{dn}{d\lambda}} + \frac{B}{f} + \frac{\nu}{2b} \frac{dn}{d\lambda}$$

This was an error in the submitted manuscript. The above formula was used in calculating the results of Table I.

Discussion with Dr. Ganz has brought out the fact that the contribution to the spectral slit width caused by the limiting resolving power of the prism, the **B** term, is not independent of the slit width, but should be multiplied by a factor which is dependent on the slit width (S). This factor is a maximum of about 0.87 when $S = 0$ as shown by P. C. Cross and E. R. Nixon,* and a minimum when $S (cm^{-1}) = \nu/b(dn/d\lambda)$. It remains constant for $S > \nu/b(dn/d\lambda)$. The use of this factor removes the discrepancy between our theoretical and experimental spectral slit widths.

* P. C. Cross and E. R. Nixon, J. Opt. Soc. Am. 34, 517-520 (1944).

Here and There

New Appointments

Effective March 1, Wallace Waterfall was appointed director of research and product development for the Celotex Corporation. He had been on leave from Celotex for three years, as a member of the scientific staff of Columbia University Division of War Research.

Arthur D. Little, Inc., industrial research organization, has announced two new appointments. Austin W. Fisher will supervise chemical engineering research, and Bruce S. Old will engage in metallurgical research.

Charles D. Coryell, who has been active in atomic research at the Clinton Laboratories, Oak Ridge, Tennessee, has been appointed professor of chemistry at the Massachusetts Institute of Technology. He will join the staff in July.

Carroll L. Wilson, executive assistant since 1942 to Dr. Vannevar Bush, director of the Office of Scientific Research and Development, has been elected vice president of National Research Corporation, Boston, Massachusetts.

Midwest Research Institute, Kansas City, Missouri, recently announced the following new appointments: Max H. Thornton and Loren E. Morey, biochemists, John D. Fulton, biological research specialist, and Howard M. Gadberry, chemist.

J. Mitchel Fain, lately a colonel with the Army Air Forces, has rejoined the staff of Foster D. Snell, Inc., as account executive in charge of technical development for a group of manufacturers of chemical specialties.

John D. Kraus, formerly of the Radio Research Laboratory at Harvard University, has been appointed associate professor in the Department of Electrical Engineering at the Ohio State University.

Effective July 29, Richard A. Beth will become professor of physics and head of the Department of Physics at Western Reserve University. Since 1940 he has been associated with the National Defense Research Committee at Princeton University, directing research on terminal ballistics.

Awards

The annual Young Author's Prize of the Electrochemical Society was presented at the Society's meeting in April to Austin E. Hardy, 25-year-old chemist in the Luminescent Materials Laboratory of the Tube Department of Radio Corporation of America. He received \$100 for his paper on "Photoconductivity of Zinc Cadmium Sulphide as Measured with the Cathode-Ray Oscillograph."

The American Russian Institute reports that the 1946 Stalin Prize for Physics was awarded to K. A. Petrzhak and G. N. Flerov for their discovery of spontaneous fission of uranium.

In recent months The Franklin Institute, Philadelphia, Pennsylvania, has presented honor award medals to the following research scientists in recognition of outstanding

work in their various fields of interest: Major General Gladeon M. Barnes, Dr. Karl Terzaghi, Dr. George C. Southworth, Dr. Henry C. Sherman, Sir Henry T. Tizard, Dr. Sanford A. Moss, Dr. Bengt Edlén, and Dr. Ira S. Bowen.

New Applied Science Department at Harvard

Harvard University announced in February the formation of a new Department of Engineering Sciences and Applied Physics. The department was organized to consolidate and strengthen the University's research and instruction in applied science and will give instruction at both undergraduate and graduate levels.

A primary educational aim of the new department will be to fill in the "blind spot" which exists between instruction afforded by departments of physics and mathematics in which primary emphasis is upon pure science, and the training afforded by mechanical, electrical, and metallurgical engineering departments in which instruction is concentrated on the technology of engineering practice. The wholesale mobilization during World War II of trained scientists and scientists-in-training taught certain lessons which are being used to guide the planning for the activities of the new department. Administrators and supervisors of many OSRD war research laboratories had an unusual opportunity to observe in action men who had been exposed to various kinds of scientific training. Most of the scientific military objectives centered around novel applications and extensions of principles and techniques which were still in the research laboratories in the late 1930's. It was an almost universal conclusion that those people succeeded best in such advanced development work who had acquired the most thorough basic training in mathematics and fundamental physics, without regard to whether their previous inclinations had veered toward pure science or engineering. The "pure scientists," in particular, were able to adapt their thinking to the requirements of engineering development with surprising facility and success. Since the postwar demands on scientific understanding are likely to become more rather than less exacting, these considerations have influenced the pronounced emphasis on basic training in the programs of study for both undergraduate and graduate degrees.

The undergraduate program will embrace Harvard's series of courses in general education as recommended in the report of the Harvard Committee on Objectives for General Education in a Free Society. It will constitute an "honors" program open only to students qualified to profit by the rigorous training in fundamental mathematics and physics. No attempt will be made to provide at the undergraduate level the various specializations which characterize the different specific branches of engineering. Emphasis will be placed instead on a strong basic training which will qualify the student to pursue graduate work and to carry on research in any of the specialized branches of applied science.

The nucleus of the new Department of Engineering Sciences and Applied Physics is composed of several faculty groups which have been engaged in instruction

and research in acoustics, communications, electronics, applied mathematics, metallurgy, and mechanical engineering science. All of these staff members will bring to their new teaching duties the experience gained from extended war assignments in military research or training where there was opportunity to observe the end results of many methods of instruction in applied science. Associate Professor F. V. Hunt has been appointed chairman of the new department, whose membership will include Professors C. H. Berry and E. L. Chaffee, Associate Professors H. W. Emmons, H. R. Mimno, H. H. Aiken, and R. W. P. King, Assistant Professor L. A. Pipes, and Doctors R. W. Hickman and P. LeCorbeiller. In addition to current endowment support for this group, the University has assigned to the use of the new department the income from \$2,000,000 of the Gordon McKay Fund, with which it will be possible to make new appointments.

Among the unique facilities which will be available for research under the auspices of the Department of Engineering Sciences and Applied Physics are the Automatic Sequence Controlled Calculator presented to Harvard in 1944 by T. J. Watson, president of International Business Machines Company, the large anechoic (echo-less) chamber constructed in 1942-43 for free field sound measurements, other rooms specially treated for research in acoustics, and a supersonic wind tunnel constructed at Harvard during the war.

The scholarship program for the department will be patterned after the Harvard National Scholarships. Stipends of these scholarships are adjusted to meet the students' need and provide opportunity for a few carefully selected candidates to pursue their entire program of study without financial concern. Other student aids and teaching fellowships will also be provided.

New A.S.H.V.E. Research Facilities

The American Society of Heating and Ventilating Engineers has purchased property at 7218 Euclid Avenue, Cleveland 3, Ohio, to provide adequate research facilities for serving the entire heating, ventilating, and air conditioning industry as well as the associated industries.

Atomic Research at General Electric

General Electric Company plans an expansion in its program for the study of the basic physics of atomic power, which eventually will require attention of "a substantial part" of its research staff. Already approximately twenty scientists who formerly were engaged on the Manhattan District Project have been appointed to the laboratory since the war's end. The company plans to add another forty technical men to its research staff before the completion of a new laboratory sometime next year.

Expansion at University of Rochester

Recently the Eastman Kodak Company and the Bausch and Lomb Optical Company presented gifts totaling \$150,000 to the University of Rochester for its Institute of Optics. This will support the Institute's teaching and research work, and provides for the appointment of

Professor Brian O'Brien to the newly created post of research professor of physics and optics.

This enlarged program is in addition to plans for expanding the Department of Physics, for which \$550,000 was recently appropriated by the University. The two programs in physics and optics will be coordinated under the administrative direction of Dr. Lee A. DuBridge, chairman of the Department of Physics.

General Electric Science Fellowships

Again this summer General Electric Company will provide all-expense science fellowships at Union College for high school chemistry and physics teachers, selected from ten northeastern States. Fifty teachers will be appointed for the six-week term of study, which will begin July 7.

Electrochemical Society Officers

At the Congress of the Electrochemical Society (first since the beginning of the war) held at Birmingham, Alabama, April 11 to 13, the following new officers were elected:

President	William C. Moore, <i>New York</i>
Vice President	A. L. Ferguson, <i>University of Illinois</i>
Treasurer	W. W. Winship, <i>New York</i>
Secretary	Colin G. Fink, <i>Columbia University</i>

New Laboratory for Cluett, Peabody and Company

In April ground was broken for a new \$600,000 research laboratory which will carry on scientific research on a scale "without precedent in its industry," according to C. R. Palmer, president of Cluett, Peabody and Company. The company's enlarged research program will be carried out along broad general lines which will include chemical process studies, textile research, mechanical process research, and fundamental research into such projects as the development of new synthetic fibers. The laboratory will contain a number of pilot plants where the results of research developments will be tested in pilot plant operation before being attempted in actual commercial production.

New Quarters for National Research Corporation

Under lease arrangements with Massachusetts Institute of Technology, National Research Corporation will build new quarters on the corner of Memorial Drive and Wadsworth Street, Cambridge, on the Charles River Basin. Construction should be completed by December 15, 1946. Larger facilities are needed as a result of the company's increased research operations and broadening of research programs to include in some fields work for other organizations on a contract basis.

The Bulletin of Mathematical Biophysics

JUNE, 1946

On Oblique Growth of Trees under the Action of Winds: I. OPATOWSKI
 Further Contributions to a Probabilistic Interpretation of the Mathematical Biophysics of the Central Nervous System: N. RASHEVSKY
 A Note on the Excitation of Nerve Pathways: RICHARD RUNGE
 Tree Trunks and Branches as Optimum Mechanical Supports of the Crown: I. The Trunk: MARINUS H. M. ESSER
 Outline of a Matrix Algebra for Neural Nets: H. D. LANDAHL AND RICHARD RUNGE
 Volume 8, Number 2—The University of Chicago Press, Chicago, Illinois

New Books

General Meteorology

By HORACE R. BYERS. Pp. x+645. McGraw-Hill Book Company, Inc., New York, 1944. Price \$5.

The extensive wartime training of new weather forecasters for the Army and Navy severely taxed the university meteorology teachers. The few available textbooks tended to be either too specialized and advanced for the student's comprehension, or too elementary to teach him enough for his future job. Consequently, many professors wrote self-contained and comprehensive lecture notes at an intermediate level, some of which have recently appeared as books. Because of their origin, these books provide an unusual opportunity for the non-meteorological reader to acquire a very considerable knowledge of weather science without having to refer to numerous specialized texts and articles.

General Meteorology, by Professor Horace R. Byers of the University of Chicago, appears to be the type of book described above, although it is partly based on the same author's 1937 book, *Synoptic and Aeronautical Meteorology*. The great part (in general, the last two-thirds) of *General Meteorology* is devoted to synoptic meteorology—that is, the description of the earth's air masses and their interactions, the use of weather charts in describing and forecasting the weather, the general circulation, and such special topics as tropical cyclones. Also included in this part are five valuable chapters of particular importance to aeronautical meteorologists; these discuss precipitation, fog, icing, thunderstorms, and turbulence. The author is one of the nation's leading synoptic meteorologists, and the chapters on this subject are a gold mine of both classical and recent information and theories. This descriptive material and its presentation are alone quite enough to make *General Meteorology* an excellent book. The chapter on fog seemed especially good to the reviewer.

There are abundant references to the literature, and 300 well-drawn figures and maps illustrate the text. Three appendices give code tables, unit conversions, and a number of graphs useful to forecasters. The typography is excellent, and the reviewer found almost no misprints.

Professor Byers states in the preface that at present "More and more dynamic meteorology and synoptic meteorology lose their separate identities, and a general broad treatment of the subject becomes necessary." So we find that one-third (in general, the beginning) of the book is devoted primarily to such topics as atmospheric radiation balance, meteorological instruments and observations, thermodynamics, and the hydrodynamics of the atmosphere. The chapters on instruments are very good and as up-to-date as military censorship could permit in 1944. The dynamics material chosen for treatment is generally well-selected, and one topic (turbulence) is nicely blended with the related observational information. Some of the material, like Elsasser's radiation chart, is believed to appear here for the first time in a general textbook, and it is very welcome. One cannot feel, however, that the author

has made much progress toward the truly difficult goal of integrating dynamic and synoptic meteorology in a "general broad treatment."

A more serious deficiency of the dynamics chapters results from the author's apparently limited skill or interest in writing mathematics and physics. In these chapters, one finds too often that the definitions are loose, that the discussion is vapid, and that the mathematical derivations are clumsy. Worst of all, the physics is occasionally wrong (see, for example, the derivation of the Coriolis force at the top of page 192). While the inclusion of these chapters serves a very useful purpose, one can only wish that they were up to the standard of the rest of the book.

The book is especially recommended to persons interested in modern developments in descriptive and synoptic meteorology. It will also be well liked by many readers who want a general survey of meteorology complete in one volume.

A list of chapter titles follows: The Sun and the Earth; The Nature of Radiation; The Heat Balance of the Atmosphere; The Distribution of Temperature; Observations and Station Instruments; Upper-air and Cloud Observations; Thermodynamics and Statics; Horizontal Motion in the Atmosphere—the Winds; The General Circulation; Factors Determining Air Mass Structure; Air Masses of North America; Air Masses of Europe, Asia and the Southern Hemisphere; Frontogenesis and Frontal Structure; Fronts and Cyclones; Technique of Synoptic-chart Analysis; Highs, Lows and the Upper Air; Dynamic Effects in the Troposphere; Tropical Cyclones; Forecasting Technique; Condensation and Precipitation; Fog; The Formation of Ice on Aircraft; Thunderstorms and Related Phenomena; Atmospheric Turbulence and the Wind Structure Near the Surface of the Earth.

GEORGE E. FORSYTHE
Seattle, Washington

Elementary Statistics

By HYMAN LEVY AND E. E. PREIDEL. Pp. 184, Figs. 29, 5½×7½ in. The Ronald Press Company, New York, 1945. Price \$2.25.

This little book presents in simple and interesting fashion some of the fundamental concepts of statistics as taught by the authors at the Imperial College of Science, London. It appears to have been designed for students at about the sophomore level in a typical American engineering curriculum; the mathematics employed is for the most part algebra, although no attempt is made to avoid simple differentiation and integration where they can be used effectively.

The subjects covered are: types of frequency distributions; the arithmetic mean; measurement of dispersion; the accuracy of the mean; related measurements, correlation, and the correlation coefficient; elementary probability; the Gaussian "normal" distribution, and some of its properties; the binomial and Poisson distributions; and the elements of statistical quality control. The text is enlivened by many examples and a list of problems follows each of the fifteen chapters with answers at the back of the book.

The authors begin by making a distinction between a problem in which it is known that a true value exists although experimental difficulties prevent its determination, and one in which measurement presents no difficulty, but random variations in the quantity measured must be expected. As an example of each they cite "The current that will circulate in a wire of given size and material when connected to the two poles of a battery whose voltage difference is one volt," and "The number of visitors entering a museum per day." The practicability of such a differentiation is questionable. The existence of a true value of any physical property is debatable. In addition, it is doubtful that the system of causes of variability of current in an electric circuit is any less complex than that governing the number of museum visitors. The apparent difference in the two examples is caused by the great difference in "scale" of the units of measurement compared with the smallest integral units of the things measured. Unquestionably some data exhibit some of the characteristics of both classes. Thus it appears to the reviewer that this classification lacks a clear dividing line and is open to question, but it may be that, in questioning, the reader is led to an appreciation of the nature of data amenable to statistical treatment, which is, after all, the authors' purpose. Several minor typographical errors were found, particularly in exponents, but all are obvious mistakes.

Otherwise, this book should appeal to engineers and engineering students. Most American statistics texts fall into one of two classes: those designed for courses offered to sociology, economics, and education students, and those intended for mathematicians. The first have avoided all but the simplest algebra and have consequently been wordy with few developments or proofs. The others have been discouraging to the applied scientist, particularly if he is studying on his own. The book reviewed here is very readable and furnishes sufficient mathematical development to allow the student to use the results with confidence. It is recommended as an introduction to more advanced texts, and, as such, should contribute to the growing use of statistical methods.

D. K. WRIGHT, JR.
Case School of Applied Science

Engineering Mathematics

BY HARRY SOHON. D. Van Nostrand Company, Inc., New York, 1944.

The increasing use of analytical methods in engineering problems has been reflected in recent years in a rapid growth of textbook literature in applied mathematics. Among the contributions published during the war years is the book under review, a compact treatment of the mathematical background in algebra, differential equations and vector analysis, useful to the engineer. As the author states, this book "is intended to strengthen the student in algebra and to provide him with certain mathematical tools which depend on the calculus." The author is Assistant Professor of Electrical Engineering at the University of Pennsylvania.

Following a brief but clear presentation of interpolation formulae and techniques is a detailed treatment of determinants and matrices with many problems. Although the idea of a matrix and some of its properties are very clearly presented, it seems to the reviewer that an example or two indicating direct application to electric circuit theory or to a coupled vibrating system would prove highly instructive to the student engineer.

A chapter on dimensional analysis furnishes the student with an insight into a phase of applied mathematics which is altogether too often neglected in undergraduate engineering training. The numerous, complete illustrations of the methods give the reader an excellent idea of the power and utility of dimensional analysis.

Algebraic equations and their solution by exact and by approximate methods are treated in great detail in Chapters 5 and 6. It is unfortunate that Descartes' rule of signs for predicting the number of positive and negative roots of an algebraic equation is not mentioned. The inclusion of the ideas of continuity and derivative and the use of the latter in conjunction with the solution of algebraic equations lends strength to the presentation. The reviewer feels, however, that somewhat greater use of graphical aids would enhance the value of these chapters to the student.

For the most part, the discussion of ordinary differential equations and vector analysis is done in the usual manner. Gamma and Bessel's functions are introduced, and their properties are briefly discussed. Fourier series are discussed in Chapter 7, together with a numerical method of solving for the coefficients which seems to the reviewer somewhat inferior to other methods that have been devised. An interesting discussion of approximations to the Fourier coefficients concludes the chapter.

The book closes with a discussion of several problems in physics and engineering which illustrate admirably the use of the mathematical tools developed in the preceding chapters.

Taken as a whole, the book forms a compilation of mathematical background, which should enhance the powers of the serious student in his approach to the solution of engineering problems.

S. W. MCCUSKEY
Case School of Applied Science

Infrared Spectroscopy

BY R. B. BARNES, R. C. GORE, U. LIDDEL, V. Z. WILLIAMS. Pp. 236. Reinhold Publishing Corporation, New York. Price \$2.25.

The early infra-red work by Coblentz in 1905, which pointed out the characteristic spectra of each molecule, laid the foundation for the present application of infra-red spectroscopy to the qualitative and quantitative analysis of chemical compounds, which is ably presented in the book *Infrared Spectroscopy*. This recent contribution from the American Cyanamid Research Laboratories is a worthy addition to the knowledge of infra-red spectroscopy because it points out the application of a relatively new physical tool in solving analytical problems in a large chemical industry. Unquestionably the book fulfills a long-felt need

for a readable text in this rapidly expanding field, which would appeal to chemists as well as physicists.

The material on pages 1 to 113 appeared first in *Industrial and Engineering Chemistry (Analytical Edition)*, Volume 15, No. 11, Pages 659 to 709 (1943) as an article by the present authors entitled "Infrared Spectroscopy, Industrial Applications." For formal publication a bibliography of 2701 references was added under an author and subject index on pages 114 to 236. This bibliography is a unique and highly useful section both to those now engaged in and about to enter the field of infra-red spectroscopy.

Included, also, is an extensive library of transmission curves on 363 different organic compounds with accompanying empirical formulae and alphabetical indices to these reference curves. Each of the spectral curves is mapped between 1000 and 2000 cm^{-1} in a space approximately 3 by 1.5 inches. While a few curves are given to 775 cm^{-1} , it is regretted that the frequency range was not extended to 3500 cm^{-1} . However, it is recognized that no prism material, such as LiF or CaF_2 , was available for dispersion in this region at the time this work was done. Of special interest to physicists is the structural formula on the curve sheet of each compound.

The body of the book is divided into the following sections: (a) Qualitative Applications; (b) Quantitative Applications; (c) Experimental Equipment and Techniques; (d) Spectrometer; (e) Preparation of Samples; (f) Origin of Infrared Spectra; (g) Intensity of Absorption Bands; (h) Qualitative Analysis; (i) Quantitative Analytical Techniques; (j) Present Status of Infrared Spectroscopy in Industry. The theory of infra-red absorption and its relation to molecular structure are discussed to provide the essential background for detailed descriptions of techniques useful in analysis. Instead of presenting the mathematical theory of normal modes of vibration, the concept of normal modes in molecules is clarified by analogy with mechanical models. It helps those unfamiliar with infra-red spectroscopy to visualize the origin of the spectra.

A number of empirical correlations between bond frequencies and structural relationships are cited. For example, the characteristic $\text{C}=\text{O}$ frequencies in different anhydrides, esters, acids, ketones and aldehydes, and the $\text{C}=\text{C}$ frequencies in aliphatic and aromatic compounds are plotted for easy comparison. The quantitative analysis section, although not as detailed and specific as more recent publications, serves as an excellent introduction to this new infra-red application.

W. C. SEARS

The Firestone Tire and Rubber Company

Climatology

By BERNARD HAUWITZ AND JAMES M. AUSTIN.
Pp. 410, Figs. 38. McGraw-Hill Book Company, Inc., New York. Price \$4.50.

This book is intended to be an introduction to climatology for students of meteorology and it is assumed that the reader is familiar with the basic physical concepts of

the subject. It is doubtful if it will prove of much interest to the average reader who has not had a previous introduction to meteorology since the description of the various climates is given in terms of the dynamics of the atmosphere, of air mass types, and of frontal activity.

The book is divided into two parts, the first of which presents a general survey of the various meteorological elements, their distribution over the globe, and their seasonal and diurnal variation. The factors that determine the distribution and variation of the meteorological elements are discussed.

The first chapter deals with insolation, its variation, its absorption by the atmosphere, and the heat balance over the earth maintained by radiation. In the second chapter, the temperature and its distribution over the surface of the earth as a function of latitude and time and the modifications introduced by the effects of the land and sea are discussed. Lapse rates and the temperatures at high altitudes are also briefly treated. A discussion of the winds and some of the important factors that modify them follows in the third chapter.

In the following chapters, precipitation as a climatological element is discussed at some length, also the precipitation as a function of the latitude, and the effect of various factors of time and space upon its distribution is explained. Evaporation and condensation are likewise treated. After a brief discussion of air masses, climatic zones and types of climates are outlined and then discussed in considerable detail. The first part ends with a chapter dealing with various microclimatology.

The second part of the book is devoted to regional climatology: The special climatic characteristics of each continent and of the oceans are set forth in considerable detail. The effects of the mountain chains, bodies of inland waters, and ocean currents on each of the continents are discussed. The air masses that affect each continent, their source regions, their characteristics, and modifying influences are given. The influence of the physical characteristics of the various regions of the earth is explained completely and clearly.

It will be a useful textbook for every meteorologist and student of weather and climate. It contains thirty-eight figures, most of them maps and charts; it has also eighty-three tables of illustrative and useful climatic information. A pocket inside the back cover contains seventeen plates of climatic information.

JOHN G. ALBRIGHT

Rhode Island State College

Frontiers in Chemistry. Volume III. Advances in Nuclear Chemistry and Theoretical Organic Chemistry

Edited by ROBERT EMMETT AND OLIVER GRUMMITT.
Pp. 165. Interscience Publishers, Inc., New York, 1945. Price \$3.50.

This volume contains lectures delivered at Western Reserve University to an audience of graduate students and industrial chemists. The lecturers and their subjects are: A. S. Keston, "Isotopes and their Applications in

Biochemistry"; H. S. Taylor, "Applications of Isotopes in Catalytic Reactions at Surfaces"; H. R. Crane, "Techniques in Nuclear Physics"; L. G. S. Brooker, "Resonance and Organic Chemistry"; W. H. Rodebush, "The Hydrogen Bond and its Significance in Chemistry."

This book is expensive, considering its contents. The chapters by Taylor and Brooker represent excellent presentations of their subjects, although Brooker's chapter is somewhat deceptively titled since 46 of its 74 pages deal with the application of resonance theory to the problem of color and constitution. No doubt the other lectures were useful and interesting as lectures, but it seems an extravagance to present them in this expensive form since many more thorough textbook and reference treatments of the subjects are available.

A. L. ROBINSON
The University of Pittsburgh

Electron Optics and the Electron Microscope

By V. K. ZWORYKIN, G. A. MORTON, E. G. RAMBERG, J. HILLIER, AND A. W. VANCE. Pp. 766+xi. Wiley and Sons, Inc., New York; Chapman and Hall, Ltd., London, 1945. Price \$10.

This extensive and authoritative work has been compiled by the authors in recognition of a growing need for an instructive book on the theory, construction, and operation of the electron microscope. Their task has been made more difficult by the fact that a satisfactory presentation of electron optical theory and design requires a liberal use of mathematics that is frequently beyond the grasp of those who desire only to make intelligent use of the electron microscope as a research tool. In order not to discourage the practical electron microscopist, the authors have divided the book into two parts. The first is surprisingly non-mathematical while the second employs mathematics to a degree appropriate for methodical exposition.

Part I provides a qualitative introduction to electron optical principles and their application in numerous devices of such diverse character as the beam power tube, mass spectrograph, and cyclotron. Electron microscopes of all kinds are treated but the main emphasis is on the magnetic type which has proved the most successful. A chapter devoted to electron optics of high magnification contains much useful information for the practical electron microscopist. Discussion of lenses and their aberrations is undertaken in qualitative manner. Since special preparative techniques in electron microscopy are undergoing rapid development and improvement the description of this aspect of the subject lags, of necessity, behind present practice. A final chapter deals with applications and is profusely illustrated with electron micrographs representing research in many branches of science.

Part II contains a thorough, mathematical presentation of electron optical theory and design that is sufficiently broad to be of interest to many who are not primarily interested in electron microscopy. Chapters devoted to determination of potential distribution and trajectory tracing, and a section on ion optics, may be cited in this category. Electrostatic and magnetic lenses, electron

mirrors, and their aberrations are treated systematically and at length. The discussion is supplemented by a consideration of the increase of electron mass in high velocity beams. Finally, present knowledge concerning image formation in the electron microscope is summarized in a sizable chapter embracing such topics as electron scattering in thin objects, diffraction effects, contour phenomena, and resolving power, all of which will be of interest from the practical as well as the theoretical point of view.

Except for the application of the electron microscope to research, which is to some extent beyond the scope of the book, the contents and bibliography are apparently complete to about the end of 1944. Liberal use is made of explanatory diagrams, charts, and micrographs throughout. The almost complete elimination of mathematical expression from the first part detracts from its value and the somewhat arbitrary division of the book tends to place obstacles in the way of the reader who attempts to correlate pertinent facts on some topics. In logical development, the two parts would be in the order indicated in the title. An historical note is injected occasionally but oddly enough the authors do not find it necessary, until page 493, to mention Busch, who first demonstrated the application of the lens formula in electron optics.

By gathering and presenting this material, some of it hitherto unpublished, the authors have performed a service to the rapidly expanding number of workers in the field. This book will be found useful to the beginner and the advanced worker alike and will undoubtedly stimulate progress in electron microscopy.

C. E. HALL
Massachusetts Institute of Technology

Seeing the Invisible

By GESSNER G. HAWLEY. Pp. 195+xv, Figs. 71 (48 photos.), 13×19 cm. Alfred A. Knopf, New York, 1945. Price \$2.50.

This little book is directed primarily towards those readers who have a sort of generalized curiosity about current advances in science, but whose interest is not sufficiently specific to allow them to be exposed profitably to a more precise and comprehensive discussion of the operation and use of the electron microscope. This much is already implied in the internally inconsistent title; only a scientifically uncritical mind would fail to realize that either the word "seeing" or the word "invisible" is used in an unusual and inaccurate sense, and that the purpose of the title is not to describe the book, but to catch the eye. However, the author himself in the preface leaves no doubt as to the purpose of the book, and consequently the technically informed reader should not be disappointed to find a considerable degree of semi-explanation and an almost overwhelming number of colorful adjectives.

If we look at the book then from the point of view of the reader to whom it is directed, we find that it presents certain attractive features. At the start it makes worthy effort to explain why we continue to be on the search for the ultimately small, and why this search has been

definitely limited in the field of optical microscopy. The discussion of resolving power, while somewhat vague owing to the avoidance of technical words commonly found in small dictionaries, is well done, and probably will explain for the first time to many uninitiated readers the importance of resolution in determining the useful magnifying power of optical instruments.

There follows a description of the construction and operation of the microscope, and then, greatly to the credit of Mr. Hawley, a chapter on the limitations of electron microscopy. Among the welter of popular books on the unlimited wonders of science, this reviewer finds as a most welcome novelty a chapter devoted exclusively to pointing out that there are still some hard nuts to crack in our progress toward knowing all! There are also several pages given to the history of the instrument, about which one gets the impression that the author has tried to be internationally minded, but has not been thoroughly informed as to the enormous priority of the developments of the German microscopists. For example, there is the statement that as late as 1944 one would have to say that the microscope was of "little or no use in determining the irregularities on the surface of metals." Anyone familiar with the early work of Mahl and others in Germany will know that surface replicas were successfully studied some years before 1944.

The expected descriptions of the use of the microscope in studying surface replicas, small particles, and biological objects comprise the last chapters. The fields covered are the ones most seriously studied to date, and the coverage is good for such a small book. The illustrations suffer badly, however, from being neither the most current nor the best ones available, and from being in many cases reprints of previously published illustrations. The "intelligent layman" should not be blamed for feeling, from an inspection of the illustrations, that the electron microscope mainly produces hazy, non-informative pictures on a gigantic scale of objects which we used to photograph hazily on a much smaller scale.

ROBLEY C. WILLIAMS
University of Michigan

Synthetic Rubber from Alcohol

By A. TALALAY AND M. MAGAT. Pp. 298. Interscience Publishers, Inc., New York. Price \$5.00

This treatise deals with the Russian method of making synthetic rubber from alcohol and is based principally on a survey of the Russian literature on this timely subject. Theoretical aspects, technology, and production practice on the conversion of alcohol to butadiene and the subsequent polymerization to polybutadiene are given, which makes this a valuable book for research chemists, physicists, and chemical engineers who are working in these fields. The book is of value for these reasons in spite of its limited field which does not cover the two-step process of converting alcohol to butadiene, which is used in the United States, nor does it cover compounding of the finished polymers.

The first chapter contains a discussion of the chemistry and theoretical mechanisms involved in the Lebedev Process for the direct conversion of alcohol to butadiene by the use of a dual catalyst producing simultaneous dehydration and dehydrogenation of ethanol resulting in a commercially satisfactory yield of butadiene. Considerable theory is given regarding the process utilizing the Lebedev concepts of "a" and "b" fragments and how they combine to produce not only butadiene but the by-products which are formed in the process. Gorin's work, which led to the recirculation of aldehyde with rectified alcohol to substantially increase the yields of butadiene, is mentioned. Methods of purification, temperature effects, as well as the use of catalysts and their theory of action, are given in some detail.

The second chapter concerns the technology of the S.K. Process. This chapter is particularly well written as it develops the process from the laboratory stage straight through the pilot plant stage to the final commercial production installation. Sufficient diagrams and flow sheets are given along with the description to make this a real contribution. Temperature, catalysts, yields, by-products, purification, and methods of analytical control, as well as explosion hazards, are covered.

The third and fourth chapters of the book deal primarily with the polymerization of the butadiene to polybutadiene. The theory of the processes leading to the production of macromolecules, such as polycondensation and polymerization, are explained in some detail along the concepts of "ionic" and "free radical" type polymerization catalysts. Autopolymerization, along with polymerization in liquid and vapor phases, is gone into in some detail. Emulsion polymerization is also described. A complete description of the "rod" and "rodless" methods of sodium polymerization are given along with the yields and effects of impurities.

The fourth chapter is concerned mainly with the physiochemical properties of the polymer. Considerable information on molecular weight, and molecular weight distribution, is given along with the various methods of determining molecular weight. The study of the properties of solutions and cements includes not only a study of viscosity, molar weight, and structure, but also a discussion of adhesion of these cements.

All of the chapters of this book are written in a clear and concise manner. A resumé of the contents of the chapter is given in the first paragraph, followed by a detailed description of the subject matter. Sufficient details are given in most cases so that laboratory set-ups can be duplicated, and sufficient flow sheets are given so that plant operations can be followed through. This book will be of particular interest to research chemists and physicists working in high polymer and related fields, not only because of the valuable information and logical presentation, but because of the large number of literature citations of which there are some five hundred and eighty in number. Complete author and subject indexes are included.

R. S. HAVENHILL
St. Joseph Lead Company

Dr. W. C. Röntgen

By OTTO GLASSER. Pp. 169, Illustrated, 21×14 cm. Charles C. Thomas, Springfield, Illinois, 1945. Price \$4.50

This little volume, sponsored jointly by the American Roentgen Ray Society, the Radiological Society of North America, and the American College of Radiology, has been prepared to celebrate the one-hundredth anniversary of the birth of W. C. Röntgen, and the fiftieth of his discovery of x-rays. It gives a brief, well-illustrated account of his life and work. A considerable portion of the text is occupied by new translations of the three classic papers in which Röntgen first announced the discovery of the rays and some of his subsequent investigations of their properties. They read a little more smoothly than those made by G. F. Barker in 1896 and 1897 which can be found reprinted in part in the *American Journal of Physics* 13, 281 (1945), but a quick comparison discloses no essential difference in meaning.

The whole book is based largely on Glasser's earlier volume, *Wilhelm Conrad Röntgen and the Early History of the Roentgen Rays*, which is detailed and authoritative. Some new information is included, derived from "numerous letters, documents, and other communications obtained subsequent to that publication." A reader who is not making an intensive study of the subject, however, will gain from the smaller volume a clearer and more human picture of Röntgen, and will appreciate better the brilliance of the mind which led him, step by step, in inevitable logical sequence, through the series of experiments by which he convinced himself of the existence of a new kind of rays.

His first three papers on x-rays are models of clarity and straightforward exposition. Seldom was he led aside from the main theme by mistaking as significant an observation which today would be classed as trivial; he could seize upon an observation, select it like a magician from the worthless superficialities which threatened to conceal its true value, and devise crucial experiments from which a new property of the rays would emerge with such directness that the process of scientific discovery, to the uninitiated, was made to appear one of the simplest of man's activities.

Perhaps it was this masterly but objective simplicity of his experiments and writings which caused some less able minds to be skeptical, to be critical, and to belittle his achievements, and made Röntgen, at the same time, oversensitive to these adverse buffets of fortune. Perhaps, too, some of the misunderstandings arose from the haste with which his discovery after its first publication, was transmitted to the whole civilized world, often by individuals who could neither understand nor evaluate it. In an apparently reputable book, published in 1896, one finds, for example, Röntgen's papers catalogued under the heading *X-rays and a New Art*. Whether or not the compiler intended this as a translation of the original title, *Eine Neue Art von Strahlen*, some seeds of misinterpretation were undoubtedly sown.

In reading about Röntgen's life, scientists will find that some difficulties of their profession have not changed or

vanished in the intervening fifty years. They will understand his high hopes in moving to a new position, and his disappointment if, as sometimes happened, the equipment did not come up to expectations. They will be interested in his struggle to gain a foothold on the lowest rung of the academic ladder, an achievement that was made difficult by his somewhat unorthodox educational background. They will recognize a modern parallel in the fact that ability can very often be spelled without the letters Ph.D.

The book is strongly recommended, especially to young scientists.

THOMAS H. OSGOOD
Michigan State College

New Booklets

The *General Radio Experimenter* for February, 1946, features "The Constant Waveform Frequency Meter" Eight pages. Published by General Radio Company, 275 Massachusetts Avenue, Boston 39, Massachusetts.

The Chemical Publishing Company, Inc., 26 Court Street, Brooklyn 2, New York, has issued a new catalog of technical books of all kinds. The catalog gives the date of publication of each book as well as price, number of pages, detailed descriptions, and full table of contents. Free on request.

Wabash Corporation, Photolamp Division, 335 Carroll Street, Brooklyn 31, New York, has published the latest phototechnical data on the complete Wabash Photolamp line with detailed exposure information, in Bulletin No. 743P. Photoflash and photoflood tables for use with all black and white and color films are included, with a special section devoted to color photography with both standard and daylight-blue photoflash bulbs. Four pages. Free on request.

Miniature Precision Bearings of Keene, New Hampshire, manufacturers of miniature radial, pivot, and special bearings, precision parts and tools, have issued a six-page leaflet entitled "Ball Bearings for Precision Instruments." Free on request.

Bulletin of the Atomic Scientists of Chicago for February 15 discussed "Operation Crossroads," and contained the fourth in a series of articles on atomic energy controls: "Aerial Survey—An Aid in Mining Control." The office of this organization is in Room 6, Social Science Building, University of Chicago, 1126 East 59th Street, Chicago, Illinois.

North American Philips Company, Inc., 100 East 42 Street, New York 17, New York, has recently published a 16-page booklet entitled *Phosphors and Their Behavior in Television*. Photographs and graphs included.

Journal of Applied Physics

Volume 17, Number 7

July, 1946

A Spark Shadowgraphic Study of Body Waves in Water*

J. HOWARD McMILLEN** AND E. NEWTON HARVEY

Department of Biology, Princeton University, Princeton, New Jersey

(Received April 17, 1946)

Shadows of shock waves in water were photographed by means of a point-spark of less than one-tenth microsecond duration; the waves were formed when high velocity spheres struck the water surface. The shock wave front had the form of an elliptical arc of small eccentricity; the ellipticity was caused by the greater strength and velocity of the wave at its center. The absence-of-light band had a varying width which depended on the strength of the wave. A large number of secondary waves behind the impact wave were observed; these were produced by the vibrating spherical missile as it traveled through the water. The period of vibration was determined from the spacing of the waves for $4/32''$, $6/32''$, and $8/32''$ spheres; these agreed to within 19 percent of the periods calculated on the basis of Lamb's theory. Interference of the vibration waves was also observed when two spheres were shot into the water simultaneously. The cavity behind the missile was nearly conical in shape with the forward tip being distorted by pressure changed immediately ahead of the missile. A sphere produced a cusp at the forward tip of the cavity; cylinders and cones each produced their own

characteristic shadow pattern. When a grid was placed in front of the tank, distortions in its shadowgram were observed. These were caused by the high pressures surrounding the missile and, from the shift in the grid pattern, the direction and magnitude of the pressure gradient could be inferred. When the shock waves were reflected with the wave front nearly normal to the surface there was an indication that the stem, which is characteristic of Mach reflections, was present. All other reflections were regular. Focusing of the wave by a brass mirror was demonstrated in the spark shadowgram. Waves reflected near normal incidence from substances having a lower acoustic impedance than water, were reflected with a presence-of-light band ahead of the usual absence-of-light band. At glancing incidence this reversal occurred even when the reflecting substance had a higher acoustic impedance. Striations frequently appeared in the shadowgrams when waves were reflected. These striae were from 0.02 to 0.08 cm apart and represented a quasi-stationary state. They were believed to be formed when the reflected vibration waves passed through the incident waves.

THE spark shadowgraph is one of the most useful methods of studying transient phenomena. In this method a point-source spark of microsecond duration is employed to record photographically the shadows of moving objects. Not only is a sharp silhouette of the moving objects reproduced, but any changes of

density that might occur in the medium are also recorded. Density changes may be produced by waves of displacement, by regional alterations in temperature, and by the mixing of different liquids. The first spark shadows were produced by Dvorak¹ in 1880; heated gases from a flame and striations in optical glass were some of the shadows studied. C. V. Boys,² in 1893, photographed the shadows of fast moving bullets and the acoustic waves which accompanied the mis-

* This work has been carried out under a contract recommended by the Committee on Medical Research between the Office of Scientific Research and Development and Princeton University.

** Now in the Research Department, Naval Ordnance Laboratory, Washington, D. C.

¹ V. Dvorak, *Ann. d. Physik* 9, 502 (1880).

² C. V. Boys, *Nature* 47, 415 (1893).

siles. R. W. Wood,³ in 1900, and Foley,⁴ in 1912, obtained shadowgrams showing the reflection and refraction of the acoustic waves in air, the waves being produced by sparks. Since then many investigators have used the spark shadowgraph to study both the permanent and transient changes in fluid density.

A report recently released for publication by the U. S. Navy Department reprints several spark shadowgrams showing waves produced in water by explosions. These photographs were made by P. Libessart and appear in an article by H. Polachek and R. J. Seeger.⁵ Schlieren photographs of waves in water have also been made; these were spark excited and were photographed by O. v Schmidt.⁶ Other photographs of body waves in water, using processes similar to spark photography, have been made by: D. P. MacDougall, C. H. Messerly, and E. M. Boggs; E. C. Campbell and C. W. Wyckoff; R. W. Spitzer and R. S. Price; and A. M. Shanes.⁷

In the present investigation the spark shadowgraph was employed to study body waves in water when these waves were produced by the impact of an object on the water surface. Some of the results of this investigation have already been published.⁷ In all experiments the wave was produced by the impact of a fast moving sphere on the water surface. Many of these impact waves were strong enough to be shock waves and were observed to possess the usual characteristics of such waves. Others of these waves were either weak shock waves or simply acoustic waves, it being very difficult to distin-

guish between these two wave types when they occur in water.

1. APPARATUS AND METHOD

The illuminating spark was patterned after a spark which was developed jointly by the General Electric Company and Aberdeen Proving Ground. The spark was produced between two magnesium electrodes by discharging a condenser at 4400 volts having a $0.2 \mu\text{f}$ capacitance. The light from the spark traversed a tank of water, T , and fell on the photographic plate, PP , as shown in Fig. 1. Almost all the waves which were shadowgraphed were produced by small spheres that were shot vertically downward into the water, following the path, G , in the diagram. The spheres struck the water with velocities between 1000 and 5000 ft./sec. and their sizes varied from $1/8''$ to $1/4''$ in diameter. The largest kinetic energy possessed by these missiles was about 1500 ft. lb. or 2.03×10^{10} ergs.

The spark was triggered after the wire screen, O , was broken by the missile. The time interval between breaking the screen and ignition of the spark was controlled by an electronic delay system. Shadowgraphs of spheres moving through air with velocities of around 5000 ft./sec. showed very little blurring, indicating that the effective duration of the spark did not exceed a tenth of a microsecond.

The water tank had two Plexiglas walls which were either $1/4''$ or $1/2''$ thick, depending upon the size sphere being shot. The distance between the Plexiglas walls was $5\frac{1}{2}''$ and between the other two walls it was $14\frac{1}{2}''$. The water was $10\frac{1}{2}''$ deep. The shadowgrams were recorded on 11×14 -inch lantern slide plates of medium contrast.

The shadowgrams gave a magnification of 1.11 diameters to a central object in the tank. The spark was placed in either one of two positions. In the first, the axial ray from the spark passed through the water surface and was perpendicular to the Plexiglas walls and missile path. In the other, the axial ray was lowered so as to be 10 cm below its previous position. Because of the refraction of the rays as they passed from the water to the air, two rays originating in the tank and striking the plate would be about 18.2 percent farther apart than they would be if

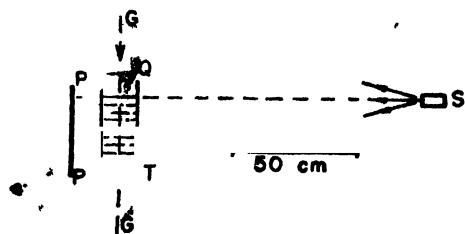


FIG. 1. Sketch of apparatus used to produce spark shadowgrams in water.

³ R. W. Wood, *Phil. Mag.* 48, 218 (1899).

⁴ A. L. Foley and W. H. Souder, *Phys. Rev.* 35, 373 (1912).

⁵ H. Polachek and R. J. Seeger, *Navy Bur. of Ord. Exp. Res. Rep. No. 14*, Aug. (1944).

⁶ O. v Schmidt, *Zeits. f. tech. Physik* 19, 554 (1938).

⁷ J. Howard McMillen, *Phys. Rev.* 68, 198 (1945).

there were no refraction. Further details concerning the apparatus are given elsewhere.⁷

2. WATER IMPACT WAVES

When a fast moving object strikes a water surface a high pressure pulse is formed. This pulse or wave travels out from the point of impact along radial lines and its front bounds a region of high pressure, high density, and high refractive index. When the impacting object transfers a large amount of momentum to the water, the wave front marks a discontinuous change in pressure and the wave is called a shock wave. The shadowgram of Fig. 2 shows such a wave; the abrupt density change across the wave front produces the light and dark bands as it is recorded in the spark shadowgram. The wave was formed by a 4/32" steel sphere striking the water with a velocity of 4840 ft./sec. and at the time the wave was shadowgraphed the peak pressure was about 141 atmospheres. At the front of the wave the refractive index was increased by 0.14 percent while the density was raised by 0.64 percent.

The width of the leading absence-of-light or dark band indicates the strength of the wave and can be used to measure the wave's peak pressure. This peak pressure, as was shown by Hilton,⁸ is proportional to the three-halves power of the band width and inversely proportional to the square root of the wave's radius of curvature. Inspection of the band width along the spherical wave front reveals⁷ that the pressure is greatest directly ahead of the sphere and that it tapers off along the arc as the surface of the water is approached. This variation in pressure is consistent with the angular distribution of the linear impulse delivered by the sphere to the water at entrance.

Shock waves in air are distinguished from acoustic or sound waves by the fact that their velocity increases with their strength or peak pressure. The same is true in water but to a lesser degree. In the spark shadowgram this dependence of wave velocity on strength is revealed by the shape of the wave front. The center portion, being stronger, travels faster than the portion near the surface so that the wave is



FIG. 2. Spark shadowgram of shock wave produced by 4/32" steel sphere. Wave had advanced 10.7 cm into the water. Debris from trigger causes weak waves in front of main wave.

not hemispherical but rather ellipsoidal. The long axis of the elliptical wave front shown in Fig. 2 is 3.4 percent larger than the short axis, showing that the average velocity of the central portion was supersonic by the same percent; the weak portion of the wave along the surface travels with the normal sound velocity. The velocity of sound in water at 15°C is 4789 ft./sec.

The shock wave in water eventually weakens and travels with a constant velocity, the acoustic velocity. The sphere on the other hand is retarded by forces proportional to the square of its velocity. Just after entering the water the sphere in Fig. 2 is subjected to a negative acceleration equal to two million g. As a result, the wave moves farther and farther away from the sphere. The sphere in Fig. 2 entered the water with a velocity of 4760 ft./sec., but at the time the shadowgram was made the velocity had dropped to 2570 ft./sec. The sphere produces a cavity which is opaque to the light from the spark and which extends all the way to the surface. Above the surface a splash is shown moving vertically upward with a velocity which, in Fig. 2, is about 1600 ft./sec. This velocity is supersonic in air and the shock wave produced by the splash is clearly visible, as are the shock waves formed in air by the sphere in its downward flight.

By measuring the band widths of shock waves formed by spheres having different diameters and impact velocities, it was found⁷ that the

⁸ W. F. Hilton, Proc. Roy. Soc. A169, 174 (1938).

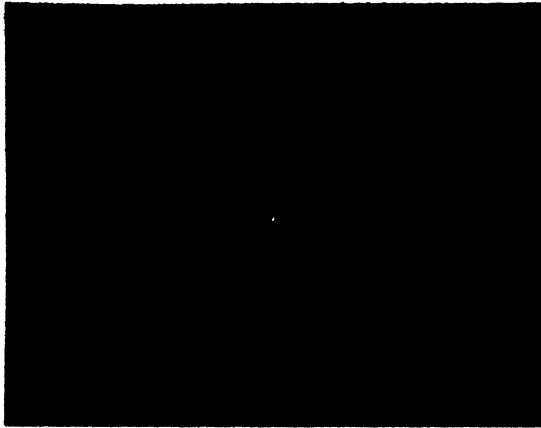


FIG. 3. Spark shadowgram showing secondary waves produced by vibrations of a $4/32''$ steel sphere. The wave has advanced 4.0 cm into the water. The bottom of the horizontal black band is the water surface, the black band being a shadow of the meniscus on the Plexiglas wall.

pressure in the wave varies as the 2.17 power of the impact velocity and is a linear function of the projection area of the sphere. The pressure which is developed at impact depends on how much water is compressed between the surface of the entering sphere and the outgoing shock wave.

3. SECONDARY WAVES FORMED BY VIBRATING MISSILE

When a missile strikes the water it is generally set into vibration and produces a series of acoustic waves in the water as it proceeds along its trajectory. These waves are found in the spark shadowgram just behind the main shock wave. Those formed by spherical missiles have been studied more extensively because of the sphere's symmetry.

A good example of secondary waves formed by a $4/32''$ steel sphere is shown in Fig. 3 where the waves completely fill the region between the vibrating sphere and the leading wave. These waves or bands are circular arcs with centers which lie between the surface of the water and the sphere, an observation which shows that they originate with the sphere as it moves forward in the water.

The bands generally form a continuous line throughout the arc covering an angular sweep of from 160° to 170° . However, it is not uncommon to find many bands disappearing or becoming indistinct along their arc. Some bands

change along the arc from sharp to diffuse or from diffuse to sharp. This change seems to take place most frequently near the angle of about 45° , as measured from the water surface.

It is interesting to note that those secondary waves or bands which lie close to the missile are not perfect arcs of a circle but are distorted. This distortion is caused by the change in the refractive index near the missile which is produced by the large pressure gradient there.

In spite of the complex nature of these bands a definite band pattern can be discerned. This pattern is more distinct directly in front of the missile; it is also easier to identify after the wave has advanced into the tank an appreciable distance. This pattern consists of a series of sharp absence-of-light bands with a broad absence-of-light band in between them. These dark bands are flanked by a narrower presence-of-light band. Figure 4 shows the band pattern produced by a $4/32''$ and $6/32''$ steel sphere. One observes that the bands of the smaller spheres are closer together. When the broad dark band is examined in the shadowgrams, as one moves away from the center along the arc, one observes that it is really two bands which only become resolved away from the center. Here it is observed to resolve itself into two narrow bands and thus the repetitive pattern is really composed of three bands, one sharp band and two bands which remain unresolved except at the side.

The separation between the sharp bands was



FIG. 4. Spark shadowgrams of the secondary waves produced by a vibrating spherical missile. The one on the left was produced by a $4/32''$ steel sphere; the one on the right by a $6/32''$ steel sphere.

measured for those waves which are in a line with the flight of the sphere. These are recorded in Table I for the 4/32", 6/32", and 8/32" steel spheres. It is observed that the band separation increases with the size of the sphere.

Since these secondary waves have centers which advance in the direction of the sphere, as consecutive bands are considered, it is clear that these waves originate with the moving sphere. This is substantiated by the observation that the waves become further apart as the sphere is slowed down. These secondary waves are produced by vibrations of the sphere; with each expansion of the sphere, an impulse is imparted to the water producing a region of increased density which travels away from the sphere as a wave. In order to calculate the time interval at which these waves are produced it is only necessary to know the band separation, S , the impact velocity of the sphere, V_0 , and the wave velocity, C . The time interval, T , is then given by $T = S/(C - V_0)$, where all measurements refer to those made in the line of flight of the sphere. The time interval between the production of successive sharp bands is recorded in the Table I. These intervals are proportional to the diameter of the spheres, showing that the band pattern is produced by oscillations of the sphere. By dividing the time interval between the sharp bands by the total number of bands enclosed by them, namely, three, one obtains the period of the vibrating sphere. The period varies from 0.97 to 1.72 microseconds.

The period of a vibrating sphere is generally expressed in terms of the time required for an elastic wave to traverse the diameter of the sphere twice. The period is given by $T = 2kD/W$ where D is the diameter of the sphere, W the velocity of the elastic wave, and k is a constant.

TABLE I. The band separation, S , of the vibration waves are given for steel spheres having three different diameters, D . The period of vibration for these spheres is also recorded and compared with Lamb's theoretical values for distortional vibrations.

D	S Band separation cm	Period of vibration microsecond	Period Lamb's theory microsecond
1/8"	0.21	0.97 ± 0.08	1.18
3/16"	0.32	1.55 ± 0.05	1.76
1/4"	0.52	1.72 ± 0.03	2.37

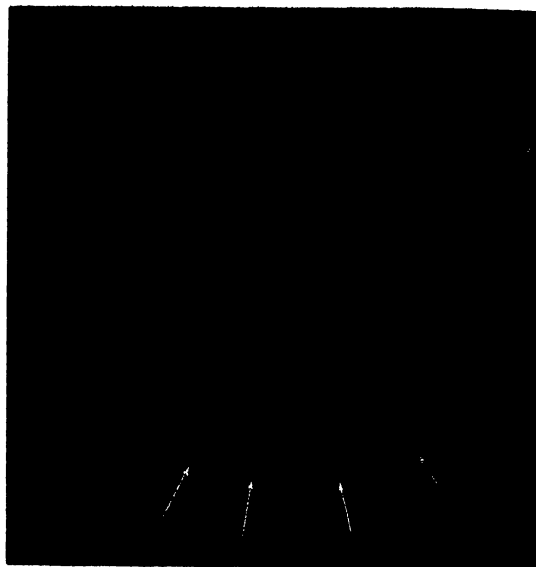


FIG. 5. Spark shadowgram of two 4/32" steel spheres shot into water simultaneously. The vibration waves produce an interference pattern; the arrows indicate the paths of reinforcement. The wave has advanced into the water 12.2 cm.

The vibrations of a sphere which give periods closest to those measured in this investigation are those in which the sphere changes shape, altering from an oblate to a prolate spheroid. For such vibrations Lamb⁹ has shown that $k = 0.595$. Using this constant and a distortion wave velocity of 3220 meters/sec., the periods of vibration for the three different sized spheres were calculated and recorded in the table. They agree to within 19 percent of the measured values.

The principal source of error here is not that of measurement but of the identification of the bands. The bands are not always clear in the photographs and further resolution may possibly show that the bands are more complex than the present experiments indicate. It is clear, however, that their origin is the vibrating sphere and that the spacing of the bands is proportional to their diameter.

A confirmatory piece of evidence regarding the nature of the vibration waves appears when two spheres are shot into the water simultaneously. Figure 5 shows two 1/8" spheres moving side by side in the water. Since both spheres are

⁹ A. E. H. Love, *The Mathematical Theory of Elasticity* (Cambridge University Press, New York, 1934).

producing waves which overlap, an interference pattern is set up. A region of reinforced waves and weakened waves is to be seen along radial paths spreading out from the spheres. These paths are indicated by arrows in the figure. The interference pattern is similar to the one observed with ripples when two oscillating prongs are moved slowly across the surface of the water. Measurement of the wave-length which produces this interference pattern, in which the distance

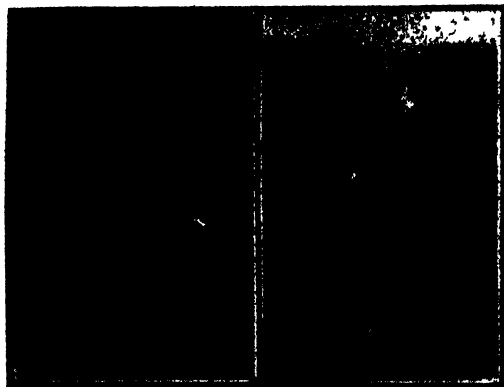


FIG. 6. Comparison of cavity shadow as produced with spark (left) and with x-rays (right). Spheres have advanced about 3.5 cm into the water, having had an impact velocity of 2700 ft./sec.

apart of the two spheres and the direction of reinforcement is used, shows that it is comparable with the spacing of the vibration waves. For example, from the interference bands in Fig. 5 one calculates that the vibration wave separation is 0.14 cm. This is to be compared with the average spacing of 0.21 cm, found with a $4/32''$ sphere. More precise determination is not feasible with the present shadowgrams.

4. SHADOW OF THE CAVITY

As the spherical missile moves through the water it tosses the water to one side, and a large cavity is produced. The pressure in the cavity is considerably less than atmospheric. For a $1/8''$ sphere traveling with a velocity of 3000 ft./sec. the cavity can have a volume which is 25,000 times that of the sphere. The cavity is nearly conical in shape and is opaque to the rays from the spark. The pressure is very great immediately ahead of the sphere, usually being measured in thousands of atmospheres, and the refractive

index of the water in this vicinity is increased. Because of these changes in refractive index the tip of the cavity produces an image on the shadowgram which is quite different from the true geometrical shadow. The tip in the shadowgram has a cusp, and it also flares out at the sides, making the shadow much wider than the cavity itself.

The true shape of the cavity was disclosed by taking x-ray shadowgrams also of microsecond duration. Since x-rays are not refracted, the true geometrical shadow was observed. Comparison with the spark shadowgrams revealed that their shadows of the cavity were as much as 250 percent wider at the tip than those produced by x-rays. For purposes of comparison, an x-ray and a spark shadowgram of a $6/32''$ sphere taken under almost identical conditions are reproduced in Fig. 6. It is clear that the center of the sphere is somewhere near the top of the cusp and the most advanced point of the sphere is level with the most advanced point of the cavity. Hence the light forming the cusp has been bent around and behind the sphere. The



FIG. 7. A shadow of an aluminum slug (with spike), taken with the slug at rest, is shown superimposed on the spark shadowgram of the slug as it is moving through the water at a high speed; the picture shows the true relationship between the slug and the cavity in the shadowgram.

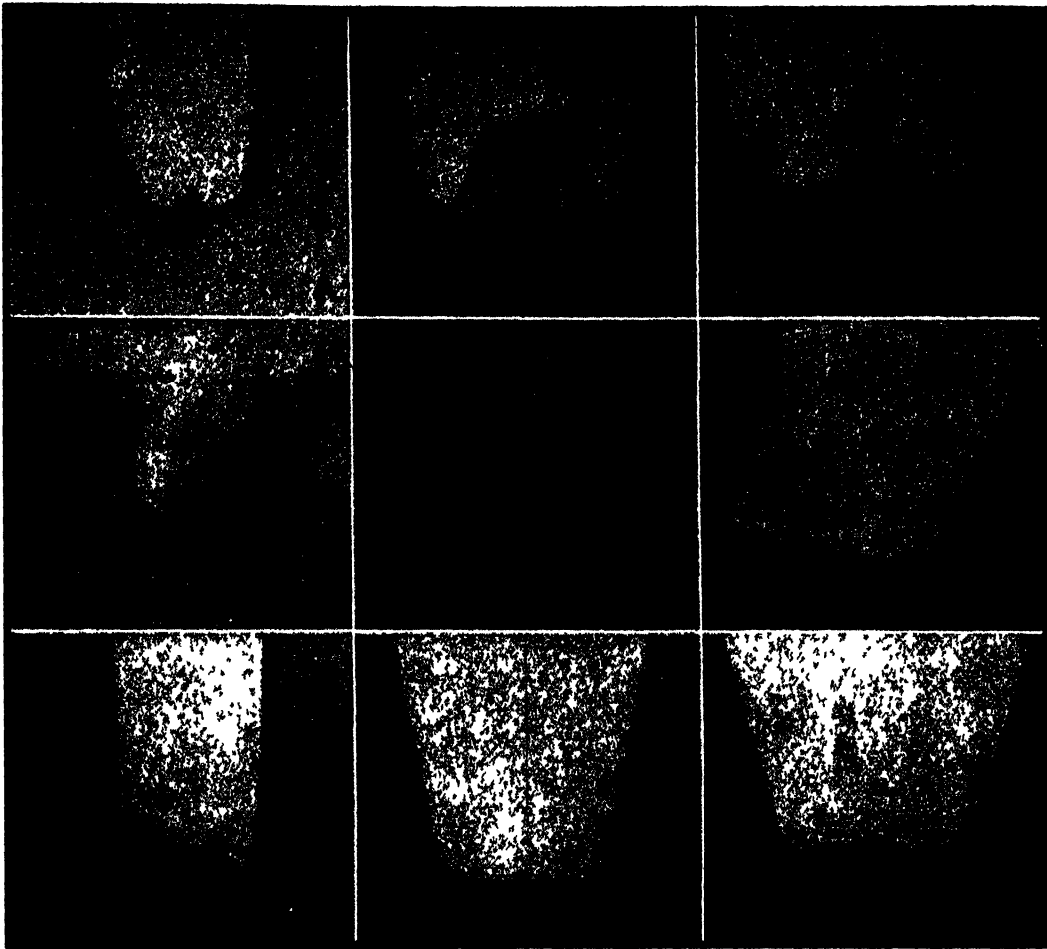


FIG. 8. Shadowgrams of the tip of the cavity as produced by various missiles. (The magnification can be determined from (G) where the cavity width is 1.0 cm.) (A) A slow $4/32''$ steel sphere at 915 ft./sec. (B) A $6/32''$ steel sphere at 1950 ft./sec. (C) A $8/32''$ steel sphere at 1710 ft./sec. (D) A $1/16''$ cylindrical dart $19/32''$ long and pointed at the ends; velocity between 1000 and 3000 ft./sec. Long axis is horizontal and perpendicular to the plane of the paper. (E) A cylindrical dart like the one in (D) with the long axis inclined about 30° with horizontal and perpendicular to the plane of the paper, velocity between 500 and 2000 ft./sec. (F) A $1/16''$ cylinder, $17/32''$ long with square ends. Velocity between 500 and 2000 ft./sec.; long axis in plane of paper. (G) A $1/16''$ cylinder, $17/32''$ long with square ends. Velocity between 500 and 2000 ft./sec. The cylinder is toppling. (H) A spin stabilized slug of aluminum with a nose having a total cone angle equal to 90° . The velocity is between 1300 and 1800 ft./sec. Total length of slug is 1.5 cm. (I) A flat-nosed slug.

position of the sphere relative to the shadowgram of the cavity was also verified by another method in which aluminum slugs were used. The aluminum slugs were machined with hemispherical ogives to which a needle point was attached. The tip of the needle showed little or no refractive effects and the hemispherical nose of the slug was located with respect to the tip of the needle by direct measurement. A spark shadowgraph of the slug at rest was superimposed on one of the slugs traveling through the water

with a velocity of 1500 ft./sec. and the composite photograph is shown in Fig. 7. The shadow of the cavity is about 38 percent wider than the cavity itself at the equatorial plane of the spherical ogive. The light forming the cusp bends in behind the spherical ogive at all points along a 50 or 60 degree arc. The length of this arc increases with the pressure in front of the sphere. The remainder of the semi-circular arc is near a low pressure region. In the low pressure region the rays are bent away from the sphere while in



FIG. 9. The distortion in the grid shows the magnitude and direction of the pressure gradient in the vicinity of a fast moving sphere. The $6/32''$ steel sphere was moving with a velocity of about 1900 ft./sec. The lines in the grid are 1 mm apart.

the high pressure region they are bent toward the sphere.

An interesting confirmation of this explanation for the large shadow of the cavity tip occurs when two spheres are shot into the water simultaneously. The flow velocity between the spheres is increased by the proximity of the two spheres, resulting in an even greater reduction in pressure on one side of each sphere. This is shown in Fig. 5 where the shadowgram shows a broadening of that lobe of the cavity tip which is closest to the neighboring sphere.

The pressures acting on the spheres are quite high. The force in dynes which acts to retard the sphere equals $0.15AV^2$, where V is the velocity in ft./sec., and A the projection area of the sphere in cm^2 . If one assumes that this force is spread out over the entire projection area of the sphere, an assumption which gives the largest possible area, then the average pressure in dynes/ cm^2 is equal to $0.15V^2$. For a sphere moving with a velocity of 4800 ft./sec. the average pressure is 3290 atmospheres, and a sphere having a velocity of 1200 ft./sec. has an average pressure of 206 atmospheres. The pressures are large enough to produce considerable deviation of the rays; the pressure of 3290 atmospheres will produce a change in the index of refraction equal to 3.3 percent.

The cusp at the tip of the cavity is more pronounced when fast spheres are used. In Fig. 8 are shown several of these cusps. The one for the $8/32''$ sphere, Fig. 8C, shows that the light has been so refracted as to focus into a straight vertical line behind the sphere. A slow $4/32''$ sphere traveling at 915 ft./sec. produces a pin-point of light instead of a cusp, as shown in Fig. 8A. The interesting cavities shown in Fig. 8D and Fig. 8E are of steel darts viewed endwise.

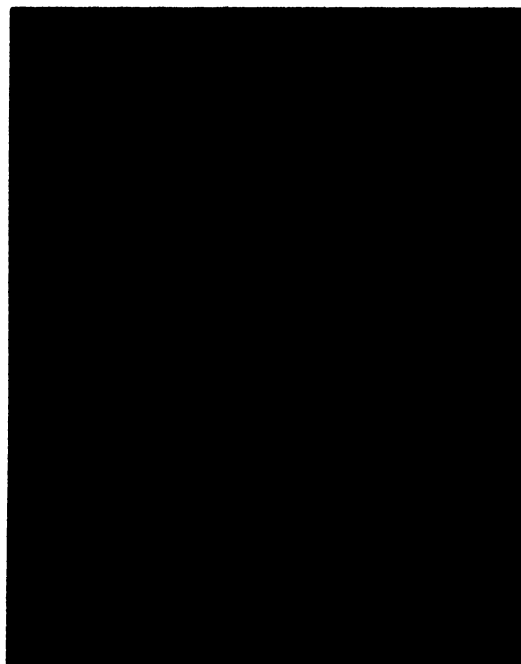


FIG. 10. A double spark shadowgraph, taken before and during the flight of a sphere through water, is shown when a one-millimeter point-grid is used. The $6/32''$ steel sphere was moving with a velocity of about 2000 ft./sec. The white streak to the right of the sphere is a blemish.

In one of these the refraction effects cause the tip to be wider than any other part of the cavity. A cylinder traveling sideways produces a shadow with bulbs at the end of the cylinder, as shown in Fig. 8F. The cylinder of Fig. 8G is toppling in a plane which makes an angle with the rays from the spark. An aluminum slug with a 90° conical nose produces the symmetrical shadow of Fig. 8H, while a cylindrical slug with a flat end gives the shadow shown in Fig. 8I. By inspecting these shadowgrams one can determine in a general way where the points of greatest pressure are located.

5. PRESSURES BETWEEN THE CAVITY AND IMPACT WAVE

The spark shadowgraph can be used to measure pressures in other regions as well as those appearing in the shock wave. For this purpose a grid was placed between the tank and the spark. The grid was made by scratching lines one millimeter apart on a sheet of Plexiglas; it was placed on the outside wall of the tank. The pattern of the grid around the missile is distorted by the changes in pressure as is shown in the shadowgram of Fig. 9. The rays above a line running perpendicular to the axis of the cavity and tangent to the two lobes of the cusp are bent downward, while those below are bent upward. The distorted pattern is also symmetrical about the axis of the cavity. A disturbance in the pattern is observed to extend to at least ten sphere diameters in front of and to the side of the missile.

When a grid made up of a series of opaque dots is used, a better idea of the distribution of pressure is obtained. This grid was made by

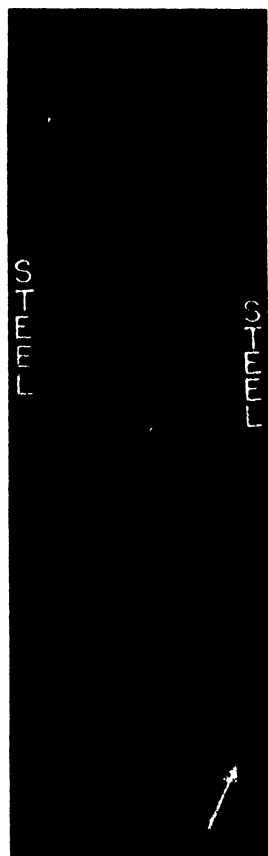


FIG. 11. Wave originating at the surface is shown being reflected from the steel plates on both sides. The arrow points to what appears to be a Mach reflection. The wave is 13.5 cm below the surface.



FIG. 12. A concave surface of brass mirror (dotted line shows surface in profile) is shown focusing the wave along *R*. Wave *P* originates at *O*, center of mirror at *C*. Concave surface of radius 8 cm was cut out of 1"×5"×5" block. Transmitted wave at *T*.

indenting a Plexiglas plate with a needle, the indentations being aligned in rows and columns one millimeter apart. The shadow cast by the indentation consists of an opaque part and a bright speck of concentrated light; the latter is black in the print. The result of recording a before- and during-the-shot shadowgraph on the same plate is shown in Fig. 10. The double image of the black dots very clearly shows the direction of the displaced ray and the orientation of the lines of the pressure gradient. Because the displacement of the rays occurs in planes which are perpendicular to the rays and which are in front of and behind that of the bullet, the pressure gradient map may be regarded as a sort of average map for all these planes. The gradient lines in front of the sphere appear to be nearly straight and to spread out radially from the nose of the spherical missile. Those behind the missile are curved, the curvature being greatest near the sphere. From measurements on the displacement of the rays it is possible to obtain values for the large pressures near the sphere. As yet these calculations have not been made because of their length; the distorted grid patterns have only been used in a qualitative way to show the direction of the pressure gradient and to indicate its magnitude.

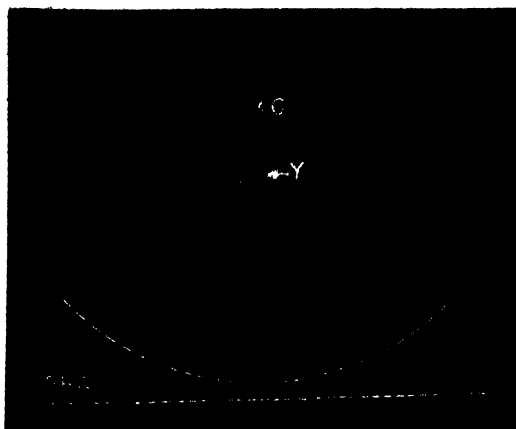


FIG. 13. Concave brass surface with center at C and radius of 8 cm is shown focusing wave at Y . Vibration waves are shown converging and diverging at the focus. Block is $1 \times 5 \times 5$ inches.

6. REFLECTION OF WAVES

Mach Reflection

Shock waves traveling through a compressible medium such as air are known to be reflected in a manner quite different from any other waves. The incident and reflected waves do not always meet at the reflecting surface but at a point just above it. In a spark shadowgram this configuration looks like a Y in which the stem is the vertical section and the incident and reflected waves are the arms of the Y. These reflections are called Mach reflections.¹⁰

Water is less compressible than air and the Mach reflections are less easily produced. Mach intersections which are similar to the Mach reflections have been observed with waves that were generated by underwater explosions.⁸ Mach reflections occur when the wave is either very strong or when the wave is reflected near glancing incidence. Since impact waves are much weaker than explosive waves, Mach reflections, if they are to appear at all, would appear at very small angles of incidence. In Fig. 11 is shown a wave being reflected from a steel plate at a glancing angle of 9.5° . The pressure in the wave is about 40 atmospheres. A stem, perpendicular to the plate, is visible although it is not nearly as well defined as those observed in air. The arrow in the picture points to the stem. For a good example of the Mach Y's in air, one

is referred to Fig. 2 and Fig. 4, where the wave from the splash and the head wave of the missile intersect.

Reflection from a Spherical Surface

When the acoustic waves in water strike a concave surface, the curvature of the wave is inverted in the same way that it is for light waves and sound waves in air, thus converting a diverging wave into a converging one. In Fig. 12 is shown a diverging wave front being reflected by a brass mirror having a radius of curvature equal to 8 cm. The dotted lines in the picture show the profile of the mirror; the origin of the wave is at O and the center of curvature of the mirror is at C . The envelope of the reflected wave now converging is labeled R and the diverging incident wave P . A shadowgram of another shot triggered at a later time is shown in Fig. 13 when the reflected wave has passed through the focal point and again become diverging. The vibration waves are shown approaching and receding from the focal point, Y ; they are converging below the point and diverging above. The waves are reflected by the brass block in the same way that light waves are reflected by mirrors in a reflecting telescope and other optical equipment.

The reflection from a convex surface is demonstrated in Fig. 14 where a lenticular-shaped object of cork composition reflects the wave. The convex surface has a radius of 11 cm and the wave originates at the water surface 11.5 cm away. The reflected wave has its center about 3.8 cm below the cork reflector; it is between the principle focus of the mirror and the mirror surface, just like optical images in convex mirrors.

7. REVERSAL

Reversal at Normal Incidence

During a study of reflected waves it was observed that some reflected waves appeared with an additional presence-of-light band ahead of the wave. Reflections from air, paraffin, and cork, for example, showed the reflected wave with the presence-of-light band foremost, whereas the incident wave has an absence-of-light band in the lead. This addition to the band pattern is

¹⁰ E. Mach, Sitz. Akad. Wiss. Wien 77, 819 (1878).

shown in Fig. 15 where a wave reflected from a block of paraffin is shadowgraphed. The wave was formed at the water surface, 10 cm above the block, and has traveled about 4 cm after reflection. The incident wave can be seen at the side. The missile has not yet hit the paraffin although it has perforated the reflected wave. This new light band appears to be about as wide as the light band which regularly follows the black band. However, in most of the reflections from other media which also produce a leading light band, the bands are broad and indistinct. This is shown in Fig. 16 where a wave is reflected from a block of cork. (The cork was large grained but had been smoothed by a sander.) Soft wood also produces this type of reflected wave. Reversal by cork-composition is seen in Fig. 14 where the wave is reflected from a convex surface. Similarly when a wave is reflected from an air surface, light is thrown out ahead of the wave. In Fig. 17 the wave is reflected from the end of an air-filled cylinder having a thin rubber membrane for a cover. The membrane is somewhat obscured in the picture by the cavitation which follows the reflected wave.

Other materials such as steel, brass, Plexiglas,

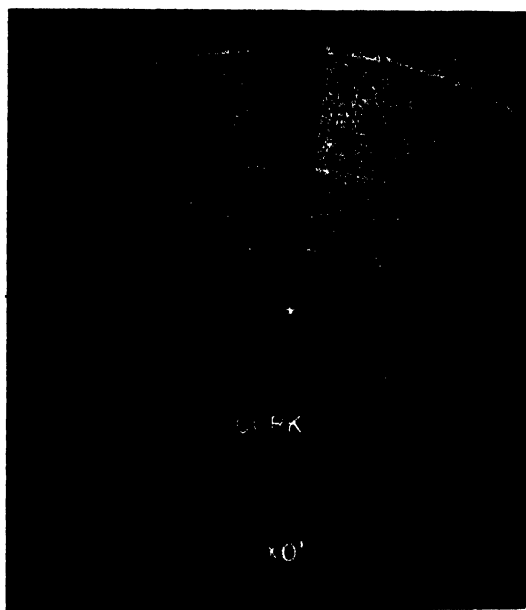


FIG. 14. Reflection from convex surface of cork showing reversal. Virtual origin of reflected wave is at O' which lies between the mirror's principle focus and its surface. Cork surface is 11.5 cm from the water surface. Note waves originating from internal reflections within cork lens.

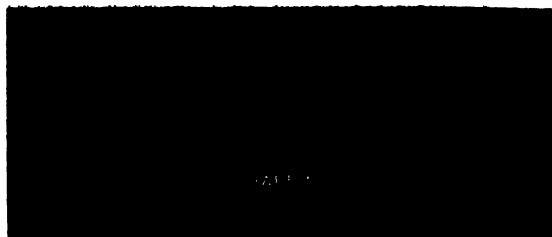


FIG. 15. Reflection from paraffin block showing reversal. Wave is 4.1 cm above block at furthest point.



FIG. 16. Reversal upon reflection from block of large-grained cork. Wave is 1.4 cm above block at furthest point.

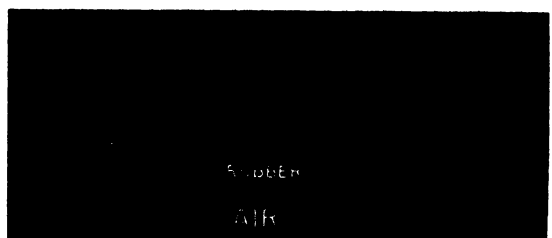


FIG. 17. Reversal of reflected wave from container of air covered with thin rubber diaphragm. Cavitation in water above rubber diaphragm causes surface to look fuzzy. Diameter of container is 12 cm, rubber is 0.028 cm thick. Wave is 3.5 cm above diaphragm at furthest point.

and lead reflect waves without any change in the arrangement of the dark and light bands. Figure 18 shows a wave being reflected from a 1/4" steel plate and Fig. 19 shows a wave reflected from a 1/2" sheet of Plexiglas. Both waves resemble the incident wave in regard to the order of bands. In the last shadowgraph there is also present a wave reflected at the lower Plexiglas surface. Since the wave in this case is being reflected by water which is less dense than Plexiglas, the wave is reversed and possesses a light band in front. Figure 20 shows a reflection from a brass plate one-half of which has been roughened, demonstrating that the smoothness



FIG. 18. Wave is shown being reflected from 1/4" steel plate. Wave has advanced 3.1 cm after reflection.

of the reflection surface does not affect the order of bands upon reflection.

Reversal of the band order takes place when the wave is reflected from a medium of lower acoustical impedance, where acoustical impedance is defined as the product of density and wave velocity. The acoustic impedance for the materials encountered in this report is recorded in Table II. All substances below water in the table have low impedances and reflect water waves with a light band in front. Likewise, waves reflected from inside those substances which are listed above water in the table also are reflected at the substance-water interface with a light band foremost. This is in accord with the accepted theory of wave reflection, namely, that when a reflection occurs in a wave going from a medium of high to low acoustical impedance there is a change in phase and the high pressure wave is reflected as a low pressure wave. Just what this change of phase does to the steep pressure front of the shock wave is not known. It is possible that only a change in the distribution of the pressure is accomplished; on the other hand a definite negative pressure region may develop. In either case the effect on the rays from the spark is to throw the beam forward producing the reversal in band order. A wave possessing a front having a pressure lower than that of the medium through which it is traveling would refract the rays from the spark so as to throw them in advance of the wave; but it is also known that there are certain distributions of positive pressure behind the front which will also refract rays so that they produce light in advance of the dark band. It is interesting to note that paraffin which possesses an acoustic impedance that is only slightly less than that of water produces a very narrow light band, while

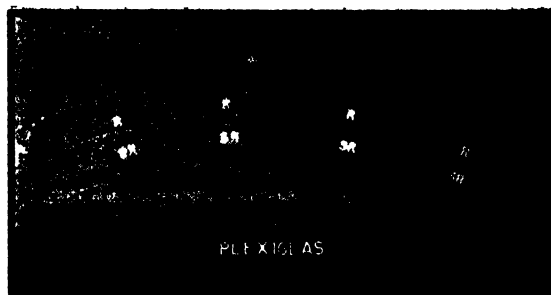


FIG. 19. Wave *R* is reflected from 1/2" Plexiglas plate without reversal. However, second wave *SR* is reflected from lower surface of Plexiglas and is reversed. Wave had advanced 4.8 cm after reflection.

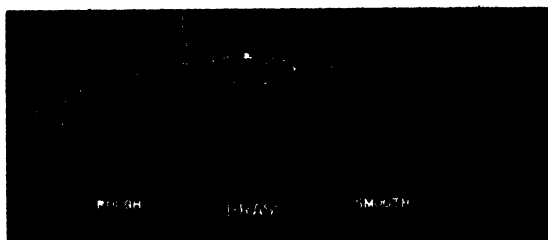


FIG. 20. Wave reflected from partially roughened brass plate showing no reversal. Left half was rough, right half smooth. Reflected wave is 5.9 cm from plate at furthest point.

other substances having a very small acoustic impedance produce wide bands.

Reversal at Glancing Incidence

When the wave front is nearly normal to the reflecting surface (the rays are then at glancing incidence), it is found that reversal occurs even when the reflecting surface has a high acoustic impedance. This is illustrated in Fig. 21 where a wave is shown being reflected from a 1/2" plate of steel. The wave front formed by rays having glancing angles of about 45° or less consists of a band of light at the front, ahead of the wave followed by a dark band. This reversal does not depend critically on the angle and

TABLE II. Acoustic impedance $\times 10^{-8}$ g cm sec.

Steel	4.67
Brass	3.60
Plexiglas	0.331
Water	0.148
Paraffin	0.126
Cork	0.099
Rubber	0.003
Air	0.000043

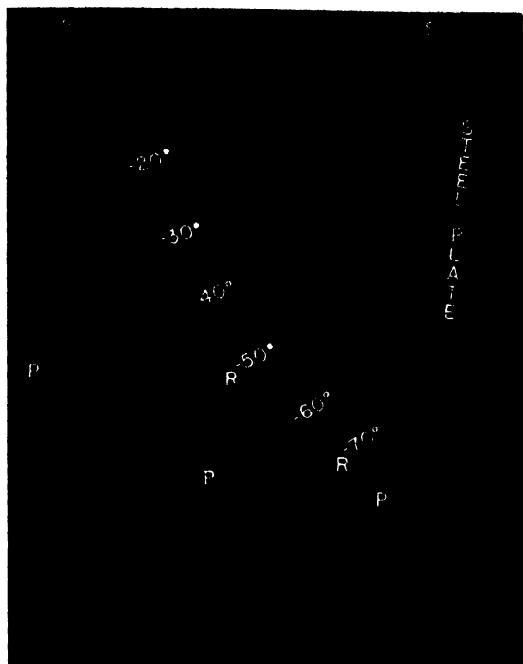


FIG. 21. Wave showing reversal upon reflection at glancing incidence from $1/2''$ steel plate. S is the surface of water, R the reflected wave, and P the incident wave of radius 13.4 cm.

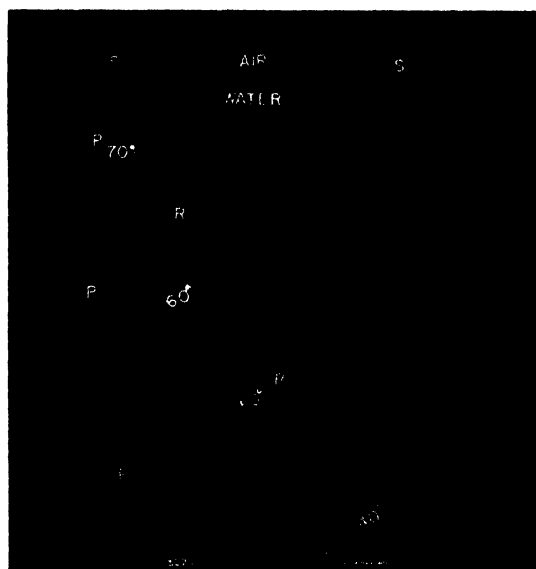


FIG. 22. Wave is shown reversed when reflected at air-water interface. S is the water surface, P the incident wave of radius of curvature 25.6 cm, and R the reflected wave. Angles of reflected rays measured with respect to the normal are also shown.

appears under various orientations of the reflecting plate. A reflection from an air surface

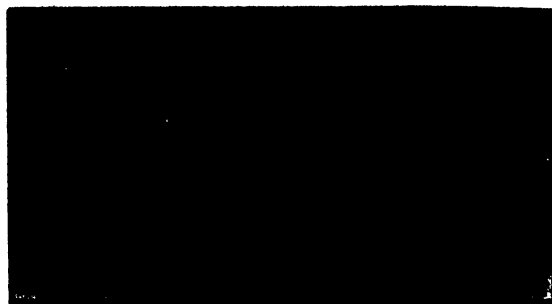


FIG. 23. Striations are shown when wave is reflected from steel plate. P is the incident wave and R the reflected. This is a detailed view of Fig. 18.

also shows reversal as can be seen in Fig. 22. It was once believed that the reversal could be attributed to the fact that the reflected wave was close behind the leading wave. When, however, an identical arrangement of wave fronts was produced by shooting two spheres into the water simultaneously, no reversal was observed; see Fig. 5.

8. STRIATIONS

A very interesting set of striations put in their appearance when the waves from vibrating spheres are reflected. The striations do not appear to move, but form a stationary pattern which persists for a very short interval. They are from 0.02 to 0.08 cm apart. Figure 23 shows these striations when the wave is reflected from $1/4''$ steel plate. The incident wave P with its vibration bands is also shown. The striae are parallel to the reflecting plate and extend throughout the region bounded by the plate and the reflected wave R . The incident and reflected rays are nearly normal to the reflecting surface in Fig. 23; when the rays are near glancing incidence, the striations produce a different pattern, as shown in Fig. 24. Here the striae follow curved paths behind the reflected wave front and are perpendicular to both the wave front and the reflecting steel plate. There are three separate waves present in this picture, for, in addition to the primary P and reflected wave R there is also the head wave H produced by the Rayleigh wave on the reflecting steel plate. In the region bounded by the head wave, reflected wave, and primary wave, the striations form a pattern of parallel straight lines.

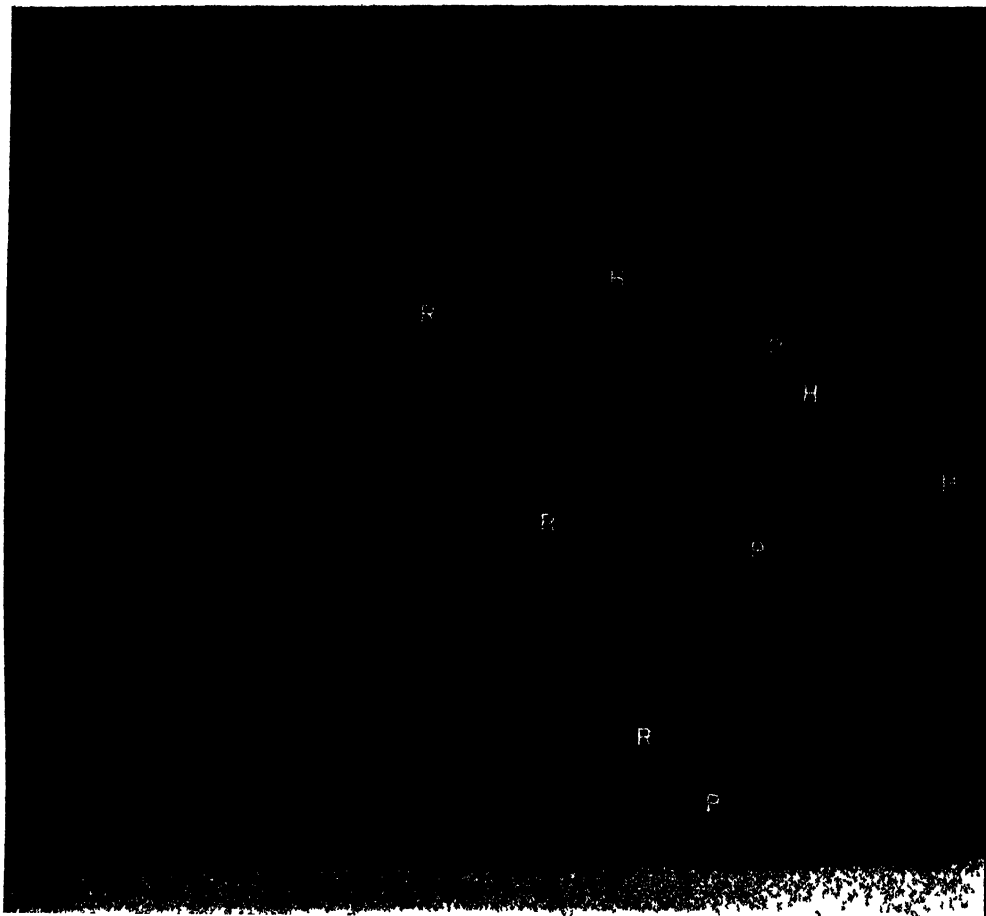


FIG. 24. Striations are shown when wave is reflected at glancing incidence from steel plate. *R* is reflected wave of radius 9.8 cm, *P* the incident, and *H* the head wave from the steel plate.

Another example of these striations is shown in Fig. 25 where a wave *RR* is shown being reflected from two lead bars, *L* and *LL*, $1/8'' \times 1''$. The bars are separated by about an inch and the diffracted waves *D* are also visible. The region traversed by the reflected wave is filled with striae. (The other objects which cast white shadows are supports that are not in the plane of the shot.)

Since the striations always appear when a reflected wave passes through a region just traversed by the incident wave, it is believed that they are produced by the secondary or vibration waves from the vibrating missile. These waves always occupy the region immediately behind the leading shock wave and intersect each other when the wave is reflected. In support of this idea one finds that the distance

between striae increases with the spacing between the vibration waves; the striae are farther

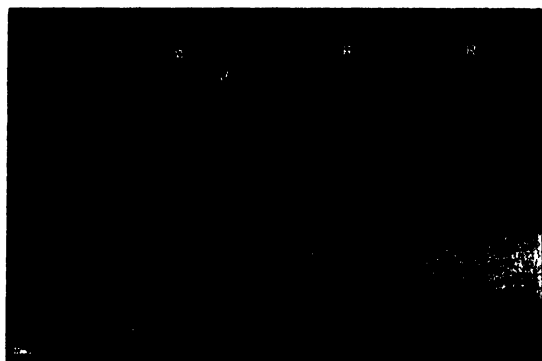


FIG. 25. Striations are shown when wave is reflected from lead bars *L* and *LL*. *D* is the diffracted wave and *R* the reflected. Note the vertical bands where the reflected waves intersect. Lead bars are $1/8''$ thick.

apart when produced by large rather than small spheres. Since these striae are oriented in a direction which is normal to a line which bisects the incident and reflected rays, it is quite possible that these striae are shadows of the wake produced by the intersection of these vibration waves. This wake may be either a region of abrupt temperature change or simply one of great turbulence. It is interesting to note that the striae occupy the same relative position as the "slip stream" does when shock waves in air intersect.

In addition to the striations in Fig. 25 a set of bands appears, where the diffracted wave from the two lead plates, *L* and *LL*, intersect. These bands resemble the optical diffraction

pattern produced by a narrow slit. While the pattern resembles an optical diffraction pattern, the pattern is more likely to be the shadow of an interference between the acoustic waves themselves. It is quite plausible that the vibration waves from the sphere, after being diffracted from the lead bars, intersect and produce these bands. The details of this wave interaction and the method by which the rays from the spark produce this effect are still obscure.

ACKNOWLEDGMENTS

The authors wish to thank Mr. Harold Towne and Mr. James Hay for their help in taking the spark shadowgrams as well as Mr. Joseph Gonzalez for his aid in preparing the prints.

Lightning Recorded by Aeroplane Propeller

J. M. BRYANT AND M. NEWMAN

Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota

(Received March 11, 1946)

A lightning stroke to the propeller blades of an aeroplane in flight left a remarkably complete record of its own characteristics. Sputtering of the discharge on the rotating blades left a clear time scale of duration of the stroke. The lightning current direction was also recorded by magnetization of steel rings inherent in the propeller mechanism.

A RECENT lightning stroke to an aeroplane in which the rotating propeller intercepted the discharge gives a measurement of the time duration of lightning at the cloud level that may be met in aircraft travel. The aeroplane was an all-metal twin-motored transport traveling at a normal cruising speed of 190 miles per hour at an altitude of 11,000 ft. A retouched photographic representation of a probable discharge path involving the aeroplane is given in Fig. 1. The propeller tip protrudes furthest below the aeroplane fuselage, thus forming the highest gradient point and hence the most probable discharge terminus of the plane when in a strong vertical electrostatic field in or between clouds.

The rotation of the propeller at 1200 r.p.m. spread out the pitting effects of the lightning discharge to a time scale, as indicated in Fig. 2, determined by the rate of rotation of the pro-

peller, and the distance of the track from the axis of rotation.

Upon examination of the propeller, shown in Fig. 1, it was found that all three blades showed burning marks owing to relatively low current values. Instead of a single stroke of short duration (40 to 100 μ sec.) usually associated with lightning discharges, the stroke lasted for at least 0.15 second. This is proven by the fact that each blade cut across the ionized path which shifted sufficiently to leave separate sweep traces so that blade No. 1 as shown in Fig. 2 was in contact with the arc at least four successive times.

The polarity of the propeller was positive to a negative cloud region as determined by the magnetic polarity of steel retaining rings about the base of the blades as shown at *R* in Fig. 3. The propeller blades and the housing are of non-magnetic material, and the steel rings were insulated



FIG. 1. Photograph showing pit marks left by a multiple lightning discharge as the rotating propeller blades swept through the stroke channel. The pit marks were retouched for clarity. The lightning channel was drawn in a probable position at the completion of the discharge process, the channel being drawn out circumferentially by the last receding propeller blade contact point.

from the blades by plastic material bushings inherent in the mechanism designed to change the propeller pitch while in flight. Currents in each blade were thus confined to flow entirely inside each ring. Each ring is split for assembly purposes and each of the three rings was found to be magnetized in the same circular direction, showing from the polarity that the current flowed through the engine shaft to each blade and thence to the cloud.

Each complete revolution of the propeller represents 0.05 second and the time scale between centers of blades represent 1/60 second. Rows of pit marks show in the forward face of each propeller blade and a more serious burning occurred on the trailing edges of the blades. One of the pit marks is shown much enlarged in Fig. 4, the depth of the pit being about one-tenth mm, the melted material being thrown outward by centrifugal force toward the tip of the blade. The whole series of events points to the prob-



FIG. 2. Photograph showing the pitting which could be caused by successive current contacting the propeller blade as it swept through the discharge channel.

ability that the plane was hit by a "multiple stroke" of lightning of relatively long duration. The nature of the discharge from the cloud entering the blades was such as not to burn the blade between pits, which could be accounted for by sudden peaks of discharge current occurring at short intervals, and thus damaging the blade in "pits" along the line of travel of the



FIG. 3. Photograph showing location of the steel rings, R , and the insulating sleeves, S , in the propeller pitch control mechanism.

arc channel. The contact time with the trailing edges was so much longer, in proportion to the ratio of the space between blades to that between pits, as to permit considerable burning.

Multiple strokes of lightning to the earth's surface have been previously recorded by photographic methods,¹ by magnetization effects in the "fulchronograph,"² and by special magnetic oscillographs. An oscillographic record is reproduced in Fig. 5 of a long duration discharge with repeated peaks, as recorded by McEachron.³ However, the time scale recording method furnished by the rotating blades gives a spacing of the successive peaks in the case of this airplane stroke as of much shorter intervals, as seen in Fig. 2, these intervals being from 10 to 100 μ sec. between the pits across the face of each blade.

It is of interest that the aeroplane propeller left a record of effects of the lightning channel near the point of origin of the stroke. From its character of shorter intervals and absence of very



FIG. 4. Photographic enlargement of a typical pit mark, showing how the molten metal was thrown out radially by centrifugal force because of propeller rotation.

¹ B. F. J. Schonland and H. Collens, Proc. Roy. Soc. A152, 595 (1935).

² C. F. Wagner and G. D. McCann, Trans. A.I.E.E. 60, 497 (October, 1941).

³ K. B. McEachron, Elec. World 113, No. 6, 430 (February 10, 1940).

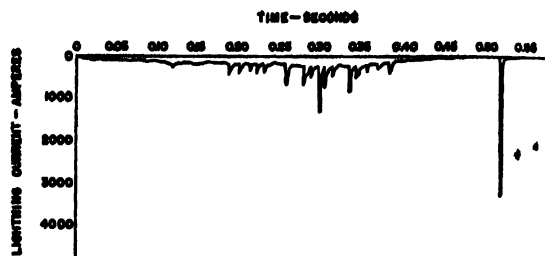


FIG. 5. Reproduction of a crater lamp oscillogram, from a paper by K. B. McEachron, showing successive peaks and continuing current in a stroke to the Empire State Building.

severe currents, it is quite likely that the stroke was from cloud to cloud rather than from cloud to ground, in which case it would be the first recorded data on such a stroke. A short cloud to cloud discharge might account for the shorter time between the consequent peaks than in those cases of strokes to ground recorded in Fig. 5. Another more likely possibility based on some preliminary calculations is that the aeroplane intercepted a discharge from a charge center, progressing upward toward the positive charges in the higher regions of the same cloud rather than to another charged cloud or to earth. Such a theory would fit the observed polarities and most probable path of the discharge, and in such a case the current peaks could correspond to sudden jumps of current in the progressive development of the lightning channel such as those observed by Schonland,⁴ for the formative stages of stroke to earth. The observed time intervals of the pitting on the propeller correspond approximately to the time intervals between the light-intensity jumps observed by Schonland. The fact that the over-all time corresponds to about ten times greater than the duration of the formative stage would fit with a case of a very long developing discharge directed through the aeroplane upwards from the cloud charge concentration rather than down to earth.

In conclusion the writers wish to express their appreciation of the cooperation of the Mid-Continent Airlines and the Northwest Airlines in the investigation of lightning strokes to aircraft.

⁴ B. F. J. Schonland, *The Lightning Discharge* (Clarendon Press, Oxford, England, 1938), p. 6.

A Microwave Dielectric Loss Measuring Technique

WILLIAM R. MACLEAN

Electrical Engineering Department, Polytechnic Institute of Brooklyn, Brooklyn, New York

(Received April 8, 1946)

Dielectric loss is measured by Q determinations of a resonator partially filled with the sample. Double sample technique is used to eliminate dominant spurious losses. Metal loss is reduced by confining field mostly to sample. Detuning for Q determination is accomplished by large movement of small rod. Detuning is calculated by action theorem.

INTRODUCTION

MEASUREMENT of the dielectric constant, ϵ , and the loss factor, d , of dielectrics at ultra-high frequencies has been described by several authors. For instance, Works, Dakin, and Boggs¹ have devised a method using a tunable reentrant resonator, and Englund² has used a transmission line method involving a plug made of the dielectric to be measured. All these methods and also the one described in this paper are based upon measurements of the Q of a resonator or other tuned circuit.

When the loss factor is very small certain difficulties arise which might be tabulated approximately as follows:

(a) Resistance losses at joints, particularly at sliding joints, can cause trouble.

(b) It often happens that the loss due to the dielectric is small compared with the loss due to the metal and, hence, one is in the unfavorable position of measuring a small quantity in the presence of a larger parasitic one.

(c) The loss due to the metal cannot be calculated due to the variability of the "roughness factor."

(d) A measurement of the loss due to the metal is often placed in some doubt if it is measured at a different frequency or if it is measured at a different internal configuration (this problem arises since a resonator without the sample does not resonate at the same frequency as with the sample).

(e) There is a certain amount of loss due to the coupling into and out of the resonator for which a correction might be necessary.

(f) Since the Q of the resonator is high in the case of small loss factors, the detuning to half-power points involves an extremely small shift in either the resonator or the oscillator. For instance, in the transmission line method it is necessary to measure a movement of only a few thousandths of an inch of the sliding plug.

(g) The half-power point of the crystal itself is at times in doubt, particularly if an assumption of a square law is made or if the crystal is calibrated at a much lower frequency.

The present method is restricted to the difficult measurement of small loss factors. If the loss factor is large, it will not work. It requires also a preliminary approximate determination of the dielectric constant, ϵ . The final measurement by the present method checks the value of ϵ and determines a small loss factor, d , by a method which minimizes most of the difficulties tabulated above. The means by which these are overcome are briefly the following:

(a) By the use of a cylindrical resonator in a transverse electric mode no current flows across the cracks.

(b) The sample is placed inside the resonator as a dielectric core which confines the field almost entirely to the dielectric. This practically eliminates the losses on the side walls. The losses on the end pieces can in principle be made relatively as small as desired by using a resonator and sample which is long enough.

(c) A calculated value for the loss on the side walls although inaccurate is sufficient to prove it negligible.

(d) The actual losses on the end pieces, whatever they may be, are eliminated in the measurement by observation taken with first one and then two identical samples.



FIG. 1. Resonator disassembled showing samples.

¹ Works, Dakin, and Boggs, A.I.E.E. Tech. Pap. 44-161, May, 1944.

² C. R. Englund, "Dielectric constants and power factors at centimeter wave-lengths," Bell Sys. Tech. J. 23, 114-129 (1944).

(e) Any loss due to the input and output coupling is also eliminated at the same time.

(f) The resonator is detuned for the Q measurements by pushing a small metal rod through the center. The detuning created by the rod is so small that a large movement may be obtained no matter how high the Q .

(g) The method of measurement includes a determination of the half-power point of the crystals at the frequency of operation.

DESCRIPTION OF APPARATUS

It is perhaps best to describe first the physical apparatus used in an experimental test of this method involving the measurement of polystyrene at 10-cm wave-length. Figures 1-3 are photographs of the equipment. In Figs. 1 and 2 can be seen a hollow metal cylinder and two identical dielectric samples and also a cover plate. Into the metal cylinder protrude two small rotatable and retractable coupling loops, one for the driving oscillator and one for the detecting crystal. To assemble the equipment one sample, for instance, is placed inside and the cover inserted. The cover is a loose sliding fit in the metal cylinder. The cover is then clamped in place with a



FIG. 2. Resonator disassembled showing samples.

clamping spider which is also visible in Figs. 1 and 2. As can be seen, the samples are bored through the center with a small hole through which moves the small metal rod also visible in the figures. Attached to the cover plate is a millimeter scale for measuring the penetration of the tuning rod into the sample. The equipment is assembled with either one or both samples inside. In Fig. 3 the equipment is shown fully assembled and ready for measurement with two

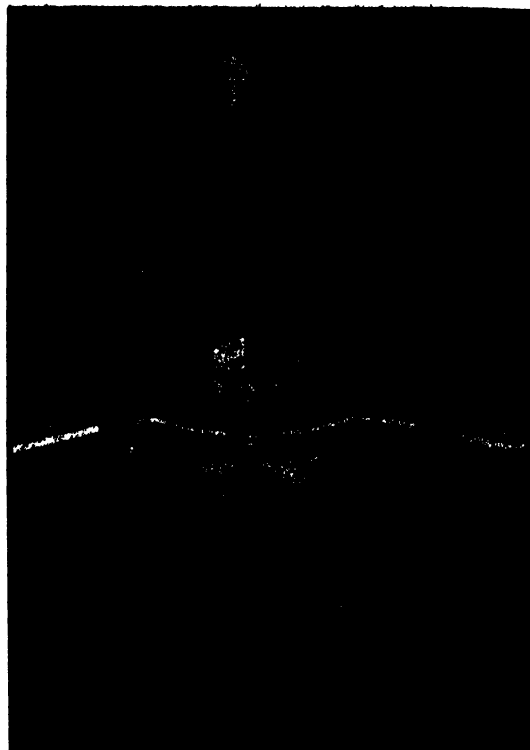


FIG. 3 Resonator assembled

samples inside. It is observed, of course, that the dielectric sample is considerably smaller than the diameter of the metal cylinder. It is an essential part of the scheme of measurement that this dielectric core confine the field to keep it away from the side walls. If the two samples are alike, the resonator will have a resonant frequency with two samples identical to that with one sample.

METHOD OF MEASUREMENT

The method of measurement is basically a Q measurement as is common to all these methods.

Assume the resonator assembled with one sample and assume a signal transmitted through the resonator by means of the two loops. To achieve this condition it will, of course, be necessary to tune the oscillator very close to the resonant frequency. It will then be found that by means of the tuning rod the resonator can be tuned to resonance with high precision. The loops are, of course, retracted until the coupling is the smallest value for satisfactory response.

Since in the mode of oscillation used the magnetic lines are vertical, the response of the crystal will vary as the loop is rotated in an approximately sinusoidal manner. Actually, the open circuit voltage generated in the crystal loop will vary exactly in a sinusoidal manner but the rotation will have a very slight effect on the impedance seen by the crystal. The weaker the coupling the less is this effect. If it were not for this impedance variation the crystal could be calibrated to its half-power point with certainty by rotating the loop 45 degrees from the position of maximum response. In actual operation with a small value of loss factor and a fairly sensitive crystal galvanometer and a driving power of the order of a watt available in the oscillator this variation in impedance will be negligible and can be so proved by repeating the calibration with two samples instead of one in the resonator; for in this case the impedance fluctuation differs by a factor of two and hence if both measurements give the same calibration this phenomenon is negligible. The crystal is now calibrated for its half-power points. With one sample in the resonator it can be tuned to resonance with the push rod, and thereafter it can be tuned to the upper and lower half-power points by advancing and retracting the push rod. Readings of the push rod penetration at the upper and lower half-power points will indicate the Q of the resonator if the characteristics of the push rod as a tuning element can be calculated. It will be shown later how this is done. If it is impossible to reach the half-power points with the push rod, then it is too small; on the other hand, if the half-power points lie too close together the push rod is too large. It will be shown later how to estimate the proper size for the push rod although this is not critical.

By this method it is therefore possible to determine the Q of the resonator when it contains first one and then two samples.

It is more convenient to calculate in terms of damping factors, d , which are the reciprocals of the Q 's. Let us suppose, for instance, that the stored energy of the resonator is given by W and the power loss is given by D . Then, as is well known:

$$d = D/W \quad (1)$$

If there are two sources of loss it is possible to consider a damping factor, d , due to each, after which the total damping factor is their sum. This is clear from (1).

In actual operation we speak of a d_1 observed with one sample and a d_2 observed with two samples. These are determined experimentally. Now suppose that d_1 consists of several parts one of which, d_0 , is due to the dielectric loss. We first seek $d_0^{(2)}$, i.e., the damping factor due to the dielectric when two pieces are used. Suppose that the same maximum field strength existed in both cases, then W and D of (1) are both doubled, hence d_0 is the same in both cases. Now let us consider d_E , i.e., the loss due to the end plates when one sample is present, and seek $d_E^{(2)}$ when two samples are present. In this case, W is doubled but D is the same, hence:

$$d_E^{(2)} = \frac{1}{2}d_E. \quad (2)$$

Consider now the loss factor d_c due to the coupling loops and suppose for the moment that in introducing a second dielectric sample we also introduced another identical pair of coupling loops with associated loads. It is obvious that then the double resonator is merely two copies of the original one and hence d_c would not be changed. But then imagine one of the pairs of coupling loops removed. This would halve the loss. Therefore the loss factors d_c obey the same law as d_E .

The losses on the tuning rod and the losses on the side walls do not behave like (2) but it turns out that they can be made negligible as will be shown later.

Neglecting then the tuning rod and side wall loss we can express d_1 and d_2 in terms of d_0 and d_x which latter represents the sum of end plate and coupling losses. These would be related by

$$\begin{aligned} d_1 &= d_0 + d_x, \\ d_2 &= d_0 + \frac{1}{2}d_x, \end{aligned} \quad (3)$$

which can immediately be solved as:

$$d_0 = 2d_2 - d_1. \quad (4)$$

We now have by a very simple formula the loss due to the dielectric in terms of the two measurable loss factors. However d_0 is the loss due to the dielectric but it is not the loss factor of the dielectric. These differ because the field is

not entirely confined to the dielectric. However a simple correction factor can be calculated as will be shown later.

DESIGN

The problem of designing the equipment will be posed as follows. The dielectric constant, ϵ , and the vacuum wave-length, λ , at which the measurement is to be made are given. It is first of all necessary to calculate the diameter, $2a$, of the sample, and its length, l . It is then necessary to calculate the diameter, $2b$, of the resonator. This can be so large that it has no appreciable effect on the frequency of resonance. That means the frequency should be practically the same as for a resonator of infinite diameter. This is readily achieved since the field is confined almost entirely to the dielectric. In addition the diameter of the resonator must be so large that the loss factor due to the side wall is negligible. It is then necessary to calculate the approximate size of the tuning rod and to calculate its calibration curve as a tuning element. Then the factor relating the dielectric loss, d , with the d_0 of (4) is calculated. One should then calculate approximately the loss due to the tuning rod.

It is not necessary to construct a new resonator for every measurement since the diameter is not critical as long as it is big enough and since the lid slides inside and adjusts itself to the sample length. As soon as these quantities mentioned above are known one can immediately proceed to the measurements just described. The rest of this article will be devoted to the calculation in convenient form of the necessary design quantities. They will be given as functions of ϵ and λ so that they can be used directly for this purpose. It turns out that there is one arbitrary degree of freedom in the evaluation of these quantities and the curves given at the end of this article are for a particular choice of this parameter, which merely determines whether the resonator should be short and fat or long and thin.

SIZE OF SAMPLE

Since it is planned to construct a resonator so that the outer wall has a negligible effect, calculations are first made assuming this radius, b , to be infinite. That is, the system is taken to be a

dielectric wave guide bounded on two ends by infinite conducting planes. A discussion of such a configuration is found for instance in Schelkunoff.³ For the present application the TE_{01} mode is desired.

In cylindrical spaces it can be shown that the complete transverse electric field can be derived from a Hertz vector, $\mathbf{Q}$, having only an axial component, Q . For this component the desired solution of the homogeneous wave equation is given by:

$$\begin{aligned} Q &= 1 \sin(\gamma Z) J_0(\rho r) && \dots \text{inside dielectric core,} \\ Q &= A \sin(\gamma Z) K_0(\alpha r) && \dots \text{outside dielectric core.} \end{aligned} \quad (5)$$

J_0 is the Bessel function and K_0 is that modified Bessel function which behaves as a negative exponential at large arguments. In (5) the intensity factor has been made unity inside the dielectric and A outside dielectric. γ is the axial propagation factor and ρ and α are the radial propagation factors related to the phase factor, β , by:

$$\begin{aligned} \gamma^2 + \rho^2 &= \epsilon_r \beta^2, \\ \gamma^2 - \alpha^2 &= \beta^2, \end{aligned} \quad (6)$$

where ϵ_r is the relative dielectric constant of the material.⁴

In this case, the formulas for $\mathbf{E}$ and $\mathbf{H}$ in cylindrical components become:

$$\begin{aligned} &\text{Inside} \\ E_\theta &= -j\omega\mu\rho \sin(\gamma Z) J_1(\rho r) \\ H_r &= -\gamma\rho \cos(\gamma Z) J_1(\rho r) \\ H_z &= \rho^2 \sin(\gamma Z) J_0(\rho r) \end{aligned} \quad (7)$$

$$\begin{aligned} &\text{Outside} \\ E_\theta &= -A j\omega\mu\alpha \sin(\gamma Z) K_1(\alpha r) \\ H_r &= -A \gamma\alpha \cos(\gamma Z) K_1(\alpha r) \\ H_z &= -A \alpha^2 \sin(\gamma Z) K_0(\alpha r). \end{aligned}$$

A must then be chosen to satisfy the boundary condition on the dielectric interface. This leads to equations for A which in matrix form would be:

$$\begin{bmatrix} \rho J_1(\rho a) & -\alpha K_1(\alpha a) \\ \rho^2 J_0(\rho a) & \alpha^2 K_0(\alpha a) \end{bmatrix} \begin{bmatrix} 1 \\ A \end{bmatrix} = 0. \quad (8)$$

³ S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), p. 425 ff.

⁴ Rationalized Giorgi units are used. μ is permeability.

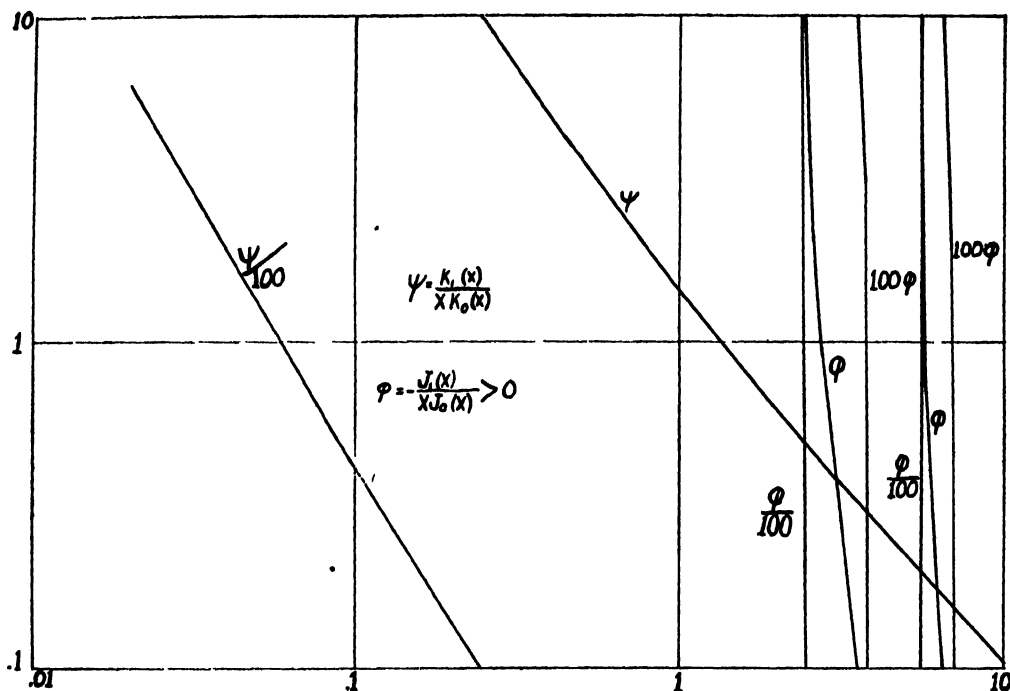


FIG. 4. Functions for symmetrical modes of dielectric wave guides.

Roots of the determinant yield the characteristic or eigenvalues of the system.

To systematize the solution somewhat we define the auxiliary quantities, x , R , by:

$$\boxed{x = \rho a} \quad (9)$$

(The equations in boxes are those pertinent to the actual design.)

$$\boxed{R = \rho / \alpha} \quad (10)$$

The determinantal equation in (8) must be solved by numerical methods, and to facilitate this we define the functions:

$$\boxed{\begin{aligned} \varphi(x) &= -\frac{J_1(x)}{x J_0(x)} \\ \psi(x) &= \frac{K_1(x)}{x K_0(x)} \end{aligned}} \quad (11)$$

which are plotted to a log-log scale as Fig. 4. In these terms the determinantal equation be-

comes:

$$\boxed{\varphi(x) = \psi(x/R)} \quad (12)$$

The design problem is usually posed in the following manner: ϵ_r and β are given and it is desired to find a , the radius of the sample and l , the length of the sample. The degree of freedom left to the experimenter which was mentioned earlier can best be used to fix the value of R . Having done this, one solves (12) for the lowest root, $x = \rho a$. Then by means of (6) and (10) the desired quantities may be expressed as:

$$\boxed{\begin{aligned} \frac{2a}{\lambda} &= \frac{x}{\pi} \frac{(1+R^2)^{1/2}}{R} \frac{1}{(\epsilon_r - 1)^{1/2}} \\ \frac{2l}{\lambda} &= \left(\frac{1+R^2}{\epsilon_r + R^2} \right)^{1/2} \end{aligned}} \quad (13)$$

In these equations λ is the wave-length in air and l is such that the sample will be $\frac{1}{2}$ wave-length long.

The value of A which is needed for later calculations now follows from (8) as:

$$A = \frac{RJ_1(x)}{K_1(x/R)} = -\frac{R^2 J_0(x)}{K_0(x/R)} \quad (14)$$

To set up convenient explicit directions for a quick application of the present method, a value for R has been chosen. I take:

$$R = 1.200$$

which gives

$$x = 2.965, \\ A = 8.66.$$

For this arbitrary choice of the degree of freedom contained in R , Eqs. (13) are plotted as a function of dielectric constant in Fig. 5. This finishes the calculation for the size of sample, since from Fig. 5 for any value of ϵ , the sample dimensions can be immediately read off in terms of the air wave-length.

CORRECTION FACTOR

It was mentioned before that the value of d_0 obtained from the measurements must be corrected to yield the true value of dielectric loss factor, d . To find this, it is necessary first to calculate the value of the stored energy, W . This involves a calculation of the integral of E^2 throughout the volume of the resonator.

In terms of evaluated Bessel integrals:

$$B_1 = \frac{x^2}{2} (J_1^2(x) - J_0(x)J_2(x)) \quad (15)$$

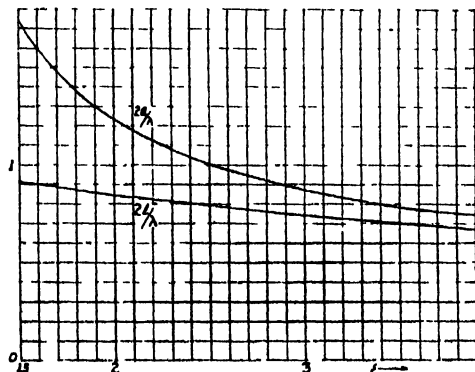


FIG. 5. Curves giving size of sample as a function of ϵ .

and

$$B_2 = \frac{1}{2} (x/R)^2 \times (K_0(x/R)K_2(x/R) - K_1^2(x/R)) \quad (16)$$

such a calculation gives:

$$W = \beta^2 \mu \pi h (\epsilon_r B_1 + A^2 B_2), \quad (17)$$

where h is the length of the resonator, i.e., l or $2l$.

Similarly the loss factor *due to* the dielectric, d_0 , can be calculated and related to the loss factor *of the* dielectric, d . The latter is:

$$d = \gamma_e / \omega \epsilon$$

where γ_e is the conductivity of the dielectric. Then the ratio d/d_0 , i.e., the correction factor mentioned above can be expressed as:

$$\frac{d}{d_0} = 1 + \frac{A^2 B_2}{\epsilon_r B_1} \quad (18)$$

For our numerical choice of $R = 1.2000$ we find:

$$B_1 = 1.075, \\ B_2 = 0.00284.$$

For these numerical values the ratio (18) is plotted as the lower curve of Fig. 6. By means of this curve the observed value d_0 can be immediately corrected to the desired value, d .

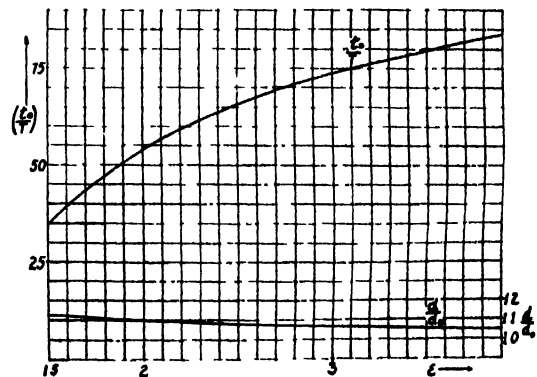


FIG. 6. Upper curve: ratio of maximum detuning to volume ratio of rod to sample. Lower curve: factor to correct for field outside sample.

SIZE OF RESONATOR

Since the lid of the resonator slides freely inside, the height of the metal cylinder is not relevant so long as it is large enough. Its diameter, $2b$, must, however, be larger than two separate numbers: namely, it must be large enough so that the change in frequency due to its presence is small, and it must also be large enough so that the loss on the side walls is negligible.

We first calculate a minimum diameter to assure only a negligible change in frequency. This calculation is carried out by imagining first an envelope placed at infinity which is subsequently brought in to the finite radius, b . At very large values of the variable radius, r , the effect of the metal envelope is to cancel $\mathbf{E}$ and the perpendicular component of $\mathbf{H}$, and also to double the parallel component of $\mathbf{H}$. This can be proved by a direct calculation, but it is also obvious from the nature of electromagnetic reflection. Such an electromagnetic field being parallel to the wall exerts thereon a pure pressure, p , whose magnitude is given by:

$$p = \frac{1}{2} \mu |H|^2.$$

The change in frequency resulting from a movement inward of the surrounding envelope can be calculated from an action theorem⁵ according to which the change in energy ΔW is related to the change in frequency $\Delta \omega$ by:

$$\Delta \omega / \omega = \Delta W / W. \quad (19)$$

(For the purposes of this calculation the damping is ignored and it is assumed that the resonator is oscillating continuously with driving forces.)

To find the change in energy of the field within the resonator when the envelope is brought from $r = \infty$ to $r = b$, we calculate the work done against the radiation pressure which can be seen to be:

$$\Delta W = 2\pi \hbar \mu \alpha^4 A^2 \int_b^\infty K_0^2(\alpha r) r dr.$$

This integral can be evaluated with sufficient accuracy by using for the modified Bessel func-

⁵W. R. MacLean, "The resonator action theorem," *Quart. App. Math.* 2, 329-335 (1945).

tion K_0 the asymptotic approximation:

$$K_0(x) = \frac{e^{-x}}{(\pi x/2)^{1/2}}. \quad (20)$$

This gives for the fractional detuning, S , the value:

$$S = \frac{\Delta \omega}{\omega} = \frac{\lambda^2 \alpha^2 A^2}{2\pi^2} \frac{e^{-2ab}}{\epsilon_r B_1 + A^2 B_2}. \quad (21)$$

However, the problem is not to determine the detuning but to determine the minimum value of b for an arbitrarily given value of detuning. Equation (21) can be re-arranged for this purpose to yield:

$$\left(\frac{b}{a}\right)_S = \frac{R}{2x} \ln \left(\frac{2}{\pi B_2 S (1 + R^2)} \times \frac{\epsilon_r - 1}{1 + (B_1/A^2 B_2) \epsilon_r} \right) \quad (22)$$

where the necessary minimum value of b as a ratio to a to satisfy the detuning condition is explicitly given.

But b must also be large enough so that the loss on the side wall is negligible. This loss, D_s , can be expressed in terms of an integral over the surface of the cylinder, and evaluated as:

$$D_s = \omega \mu \delta 2\pi \hbar b \alpha^4 A^2 K_0^2(\alpha b),$$

where δ is the depth of penetration into the metal. This value of loss yields for Q due to the shield, $Q_s = 1/d_s$:

$$Q_s = \frac{\beta^2}{2b\delta\alpha^4} \frac{\epsilon_r B_1 + A^2 B_2}{A^2 K_0^2(\alpha b)}. \quad (23)$$

The problem posed however, as was the case before, is to find a minimum value of b for a given large value of Q_s . By the use of asymptotic expression for K_0 this can be evaluated as:

$$\left(\frac{b}{a}\right)_Q = \frac{R}{2x} \ln \left[\left(\frac{\epsilon_r - 1}{1 + R^2} \right)^4 \times \frac{8A^2 \delta / \lambda}{A^2 B_2 + \epsilon_r B_1} Q_s \right] \quad (24)$$

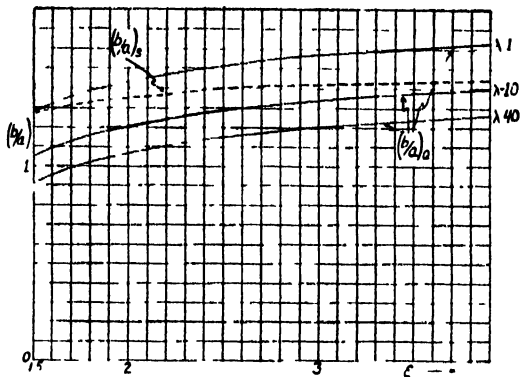


FIG. 7. Dotted curve: minimum ratio of resonator to sample diameter to assure frequency change of less than 1 percent. Solid curves: minimum ratio of resonator to sample diameter to assure side wall Q in excess of half-million plotted for various vacuum λ in cm.

For our choice of $R=1.200$, (22) and (24) have been plotted as a function of dielectric constant in Fig. 7 for the following arbitrary choice of S and Q_s :

$$S = 1 \text{ percent,} \\ Q_s = 500,000,$$

and where in (24) the value of δ for copper has been used. (24) depends explicitly on λ which is involved in δ and hence a family of the curves is necessary.

A value for b must be chosen large enough to meet the requirements of both (22) and (24).

TUNING ROD

At the center of the resonator $\mathbf{H}$ has only a z component and $\mathbf{E}$ is zero. Hence a very small axial wire all or partly immersed in the resonator does not disturb this field appreciably. Consider the circumstance of a small rod of radius, ζ , projecting into the resonator a distance z . If this rod were shrunk to zero radius and then expanded to its finite but small radius it would produce practically no change in the magnetic field H_z . However, in the process of expanding this rod, work would be done against a radiation pressure:

$$p = \frac{1}{2} |H_z|^2. \quad (25)$$

The total work done would be:

$$\Delta W = \int_0^z p \pi \zeta^2 dz. \quad (26)$$

Therefore the total work done is found by evaluating (26), and if we use the expression (17) for W and also make use of the action theorem,

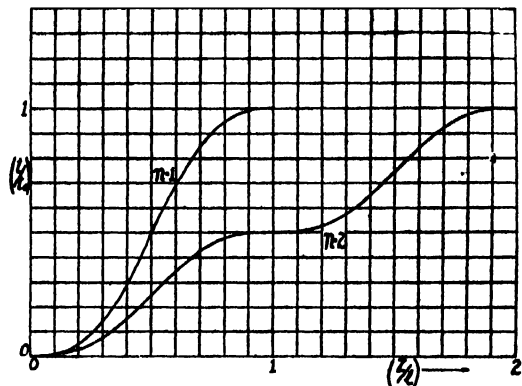


FIG. 8. Relative fractional detuning as function of fractional penetration of tuning rod for one and two samples.

we can get for the fractional detuning, t , at penetration z the value:

$$t = \frac{\Delta\omega}{\omega} = \left(\frac{1}{4\pi} \right)^2 \frac{(\rho^2 \lambda \zeta)^2}{h} \frac{\lambda_z}{\epsilon_r B_1 + A^2 B_2} \frac{2\gamma z - \sin 2\gamma z}{\epsilon_r B_1 + A^2 B_2}, \quad (27)$$

wherein λ_z is the wave-length along the axis of the resonator. In our application the resonator mode used is either TE_{011} or TE_{012} , or generically TE_{01n} , where n is the number of samples used, either one or two. Hence:

$$\lambda_z/h = 2/n. \quad (28)$$

Let us call t_0 the detuning brought about by full immersion of the rod, i.e., $z=h$. In these terms:

$$\frac{t}{t_0} = \frac{1}{n} \left[\frac{z}{l} \frac{\sin 2\pi(z/l)}{2\pi} \right], \quad (29)$$

where

$$t_0 = \left(\frac{1}{4\pi} \right)^2 \frac{(\rho^2 \lambda \zeta)^2}{\epsilon_r B_1 + A^2 B_2}.$$

(z/l) is the fractional immersion of the rod and runs from 0 to 1 or from 0 to 2 depending on the number of samples. This function is plotted in Fig. 8.

It is now convenient to introduce the ratio, T , of the volume of the fully immersed rod, to the volume of the sample. This ratio is independent of the value of n . In terms of this quantity the

total detuning becomes:

$$\frac{t_0}{T} = \frac{x^2}{4A^2B_2} \frac{R^2}{1+R^2} \frac{\epsilon_r - 1}{1 + \epsilon_r B_1/A^2B_2} \quad (30)$$

For $R=1.200$ this latter function is plotted as the upper curve of Fig. 6.

It is now also useful to check the Q due to the tuning rod. This is of course a variable quantity depending upon the immersion. However at its worst, i.e., lowest, which occurs for full immersion it can be calculated by similar processes to be:

$$Q_r = \left(\frac{2a}{\lambda}\right)^2 \frac{2\pi}{x^4} \left(\frac{a^2}{\delta\epsilon}\right) (\epsilon_r B_1 + A^2 B_2)$$

In the experimental application described, this Q_r came out to be several hundred thousand i.e., negligible.

CONCLUSION

By accepting the arbitrary choice of $R=1.200$ the curves of this article make it possible to proceed immediately to measurements by this method. Other curves could be calculated by the same process for other values of R which would yield samples of a different shape.

In the practical application of this method the writer would like to call attention to one spurious phenomenon that may occur. Because of the configuration of the field of the TE_{01n} mode the tuning rod detunes the resonator but a very small amount: of the order of one part in a thousand. However it detunes other modes radically. It is therefore conceivably possible that in pushing the rod through the resonator some other mode becomes tuned to the operating frequency. This mode will moreover probably not be confined to the dielectric and hence will couple the input and output loops very strongly. The practical result of such an occurrence would be that the galvanometer deflects sharply at one

highly critical position of the tuning rod. Unless this position coincides with a point at which a reading must be made it can merely be ignored. If it does so coincide the half-power points can readily be shifted to another region by a small change in the oscillator frequency.

The equipment described in this paper was tested by measuring the loss factor of a sample of polystyrene. To determine the reproducibility of the measurements, a series of thirteen pairs of readings of d_1 and d_2 were taken and thirteen values for d calculated. These readings yielded for the loss factor a mean value of 2.7×10^{-4} and an r.m.s. fluctuation of 0.1×10^{-4} .

As an example, the numerical values for the test sample and resonator and for one pair of measurements will be given:

Desired $\lambda = 10$ cm; assumed $\epsilon_r = 2.55$.

From Fig. 5:

$$2a = 9.9 \text{ cm; } l = 3.92 \text{ cm.}$$

Two samples cut to size. [Actual measurement of samples: $2a = 10.0$, $l = 3.92$.]

Radius of resonator chosen: $2b = 15.9$ cm. Hence $b/a = 1.59$. From Fig. 7 $S < 1$ percent and $Q_r > \frac{1}{2}$ million. Sample resonated at $\lambda = 10.03$ checking closely assumed ϵ_r .

Tuning rod diameter 0.240 cm chosen, yielding

$$T = 1/1740.$$

From Fig. 6: $t_0 = 3.91 \cdot 10^{-4}$.

Crystal calibrated to half-power points and following readings taken:

	$n=1$	$n=2$
$\frac{1}{2}$ power penetrations (cm)	$\left\{ \begin{array}{l} 1.1 \\ 2.7\frac{1}{2} \end{array} \right.$	$\left\{ \begin{array}{l} 0.9\frac{1}{2} \\ 5.8\frac{1}{2} \end{array} \right.$
fractional penetrations	$\left\{ \begin{array}{l} 0.28 \\ 0.70 \end{array} \right.$	$\left\{ \begin{array}{l} 0.24 \\ 1.49 \end{array} \right.$
t/t_0 from Fig. 8	$\left\{ \begin{array}{l} 0.12\frac{1}{2} \\ 0.85 \end{array} \right.$	$\left\{ \begin{array}{l} 0.04 \\ 0.74 \end{array} \right.$
difference	0.72 $\frac{1}{2}$	0.70
$t = d_1, d_2$	2.83×10^{-4}	2.74×10^{-4}
calculation of d_0	2.65×10^{-4}	
d from Fig. 6	2.86×10^{-4}	

Since the stability of the oscillator used was poor, and since the means for measuring the penetration of the tuning rod in the present equipment is not very accurate, it seems that better results could be obtained by better equipment.

Magnetically-Controlled Gas Discharge Tubes

R. E. B. MAKINSON, J. M. SOMERVILLE, K. RACHEL MAKINSON, AND P. THONEMANN
University of Sydney, Australia

(Received February 4, 1946)

Discharge tubes are described in which a low pressure glow discharge is initiated by means of a magnetic field pulse. A tube has been developed in which a glow discharge so triggered leads to the formation of a cathode spot on a mercury pool; this tube will hold off 20 kv until fired and will then pass a current pulse of several hundred amperes. Its use in radar modulator circuits is envisaged.

1. INTRODUCTION

IT is known¹⁻³ that the application of a magnetic field of sufficient strength can initiate a discharge between charged electrodes in certain configurations, such as coaxial cylinders, when the gas or vapor pressure is too low to permit a discharge otherwise. The magnetic field lengthens the electron trajectories so that the average number of ionizing collisions suffered by each electron is sufficient for a glow discharge to be sustained. A detailed study of such glow discharges carrying continuous currents up to several amperes between coaxial cylinders has been made by Penning,³ who has also used the same principle in the construction of a low pressure manometer and a low pressure ion gun.^{4,5}

The work described below was mostly conducted with current pulses, the peak current being of the order of a hundred amperes and the discharge normally an arc, with consequent low voltage drop between the electrodes. The main object was the development of magnetically-controlled discharge tubes which would be suitable for use as modulators in radar equipment. Both permanent gases and vapors were used, tube life in the former case being limited by clean-up of the gas. A satisfactory type of tube was developed having two coaxial cylindrical electrodes, together with a mercury pool as an auxiliary cathode* (Fig. (5a)); a glow discharge in the mercury vapor between the cylinders was

initiated by a magnetic field pulse, and bombardment of the pool by ions from this glow discharge produced an arc spot on the pool through which the main current passed. These tubes passed currents up to 200 amperes in 2-megawatt pulses of duration 1 to 10 microseconds at repetition frequencies up to 1200 c/sec. with very small power loss in the tube itself.

Although simple two-electrode concentric cylindrical tubes were not satisfactory as radar modulators it was found that in such tubes, if the glow discharge was not allowed to pass into an arc, currents up to 50 amperes could be switched off against supply voltages of a few thousand volts by removal of the magnetic field. It seems therefore that such tubes may be useful where a rapid high current or high voltage switch is needed.

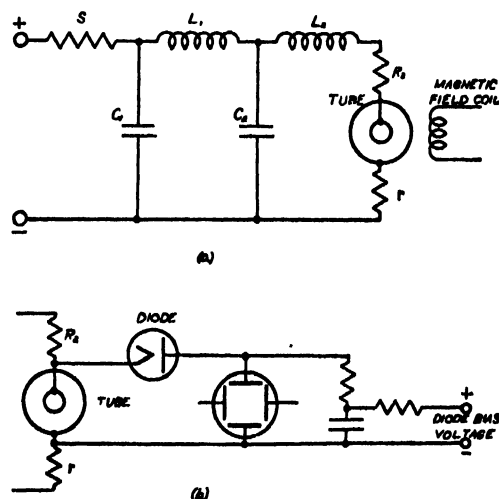


FIG. 1. (a) Showing pulse-forming line and load, R ; (b) Arrangement of biased diode to hold off H.T. from oscillograph.

¹ Strutt, Proc. Roy. Soc. 89A, 68 (1913).
² Engel und Steenbeck, *Elektrische Gasentladungen* (Verlagsbuchhandlung Julius Springer, Berlin, 1934), Vol. II, p. 236.

³ Penning, Physica 3, 873 (1936).

⁴ Penning, Physica 4, 71 (1937).

⁵ Penning and Moubis, Physica 4, 1190 (1937).

* The name "Hodectron" is proposed for tubes of this type.

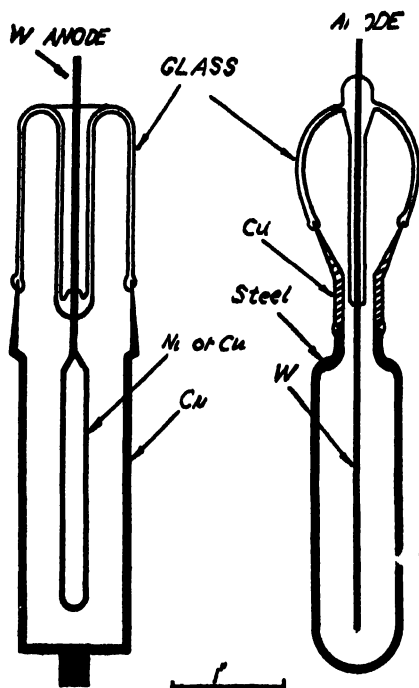


FIG. 2. Examples of two-electrode tubes.

2. CIRCUIT ARRANGEMENTS

The magnetic field was produced by a short coil coaxial with the electrodes and was applied in pulses of about $50 \mu\text{sec.}$ duration at repetition frequencies which were usually 50 or 100 per second; the peak value of the field was of the order of 500 oersted. The current in the coil was produced by charging a condenser from a d.c. source and discharging it through the coil by means of a vibrator or mercury thyratron.

The performance of the tubes was studied mainly in the circuit shown in Fig. 1 (a), which approximated that in which they would be required to operate as radar modulators. In this circuit the resistive load R corresponded to that placed on the modulator by the radar oscillator. The condenser C , forming with the inductances L_1 and L_2 an artificial line, were charged through the hold-off resistance S to between 1 and 20 kv, and were discharged through the tube and the load R on the initiation of a discharge by the magnetic field pulse. The variation of potential difference between the electrodes was studied by means of a triggered oscillograph, a diode being used as shown in Fig. 1(b) to hold off potentials greater than a few hundred volts. The oscillo-

graph was also used to study the current through the tube, represented by the voltage drop across the small resistance r .

3. GENERAL NATURE OF THE DISCHARGE IN TWO-ELECTRODE TUBES

Examples of two-electrode tubes studied are shown in Fig. 2. Essentially the electrodes were concentric cylinders; with the outer cylinder as cathode a smaller magnetic field was required to initiate the discharge, so this arrangement alone was used. Copper, aluminum, steel, and nickel were used as cathode material, copper, nickel, and tungsten for the anode. Tubes were filled with helium, argon, air, and carbon dioxide and with iodine, mercury, and water vapors.

The tubes were initially studied while still sealed to the vacuum system, so that the pressure could be controlled. The pressure at which a discharge could be initiated by a magnetic field depended on the configuration of the electrodes and the nature of the gas or vapor, but in general was of the order of 0.1 mm for permanent gases. The range of pressures for a controlled discharge in one tube was about 10 to 1.

When the applied magnetic field reached its critical value the potential drop across the tube in general fell rapidly and the condensers C discharged completely in a few microseconds. The line circuit constants determined that the current through the tube should flow as an approxi-

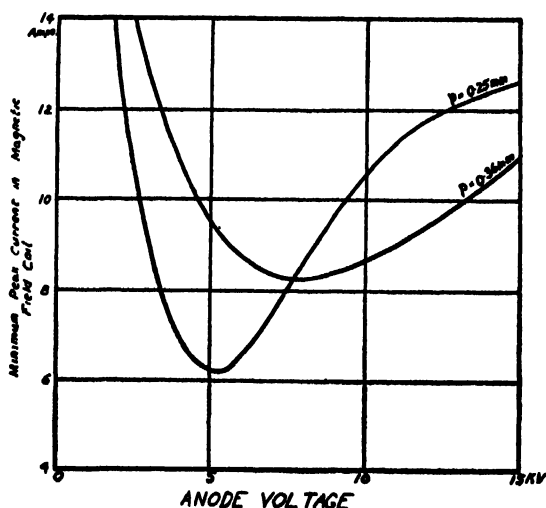


FIG. 3. Variation with anode voltage of the peak current in the field coil necessary to fire a two-electrode tube.

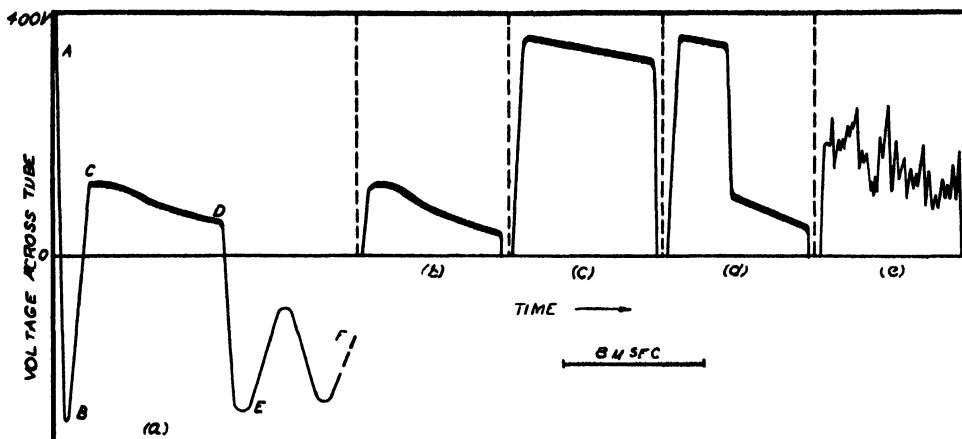


FIG. 4. Typical oscillograph traces of the voltage across the tube during discharge.

mately square pulse of a few microseconds duration; with the circuit used a small reverse current sometimes flowed after the main pulse. The currents passed were of the order of 100 amperes. The hold-off resistance S was large enough to ensure that the condensers did not charge up to breakdown potential twice during the same magnetic field pulse. The variation with anode voltage of the minimum field-coil current necessary to fire the tube is shown in Fig. 3.

Voltage Drop in Tube

When visual inspection was possible the discharge was seen to fill the whole inter-electrode space. It was usually associated with a bright white spot which moved about on the cathode and particularly favored a metal-insulator junction. Most of the current passed through this spot. When the spot was present the voltage drop across the tube was between 30 and 100 volts, in its absence about 400 volts. It was concluded that the presence of the spot meant that the discharge was an arc. Rapid alternation between arc and glow in successive discharges sometimes occurred.

Typical oscillograph traces of the voltage across the tube during the discharge are shown in Fig. 4. The initial deep valley ABC shown in Fig. 4(a) is spurious, and is caused by the rapidly falling anode potential acting through the inter-electrode capacity of the protecting diode. The portion CD shows the potential across the tube while it is conducting (the potential drop across

the diode being negligible). When the tube becomes non-conducting, the characteristics of the line produce the portion DEF of the trace. The valley ABC and the portion DEF have been omitted from Figs. 4 (b), (c), (d), and (e). Figures 4(a) and (b) are typical of arc discharges, the voltage dropping rapidly to a value of about 50 volts, which does not vary greatly either with current, over a range of 100 to 1, or with the nature and pressure of the gas in the tube. Figure 4(c) is typical of a glow discharge, the voltage usually falling to between 300 and 600 volts, although much higher values have sometimes been observed; these traces vary markedly with tube conditions and current. Figure 4(d) shows a delayed transition from glow to arc.

The arc discharge in a helium-filled tube with a copper cathode was a characteristic pink color and photographs of the spectrum showed both He and Cu lines. In tubes of this type another kind of discharge was also observed, bright green, with a voltage drop of the order of 700 volts and a spectrum showing more intense copper lines than those of the arc discharge.

The transition from the glow to the arc was favored by irregularities or discontinuities in the electrodes or the magnetic field and by an increase in the supply voltage (and consequently in the tube current). The glow-arc transition showed no marked dependence on the nature and pressure of the gas.

Owing to the low power dissipation in the tube with the arc discharge, this mode is to be

preferred. Consistent striking of the arc during the whole life of the tube was not obtained using a solid cathode. Tubes are described in Section 4 in which this difficulty was overcome by using a mercury pool as cathode.

Figure 4(e) shows an unstable type of voltage trace, frequently observed, in which the voltage was subject to rapid, apparently random fluctuations, no two consecutive traces being the same. Because of its appearance on the oscillograph such a trace has been called "grass." The height of the peaks varied from a few volts to about 200 volts. In tubes of the mercury pool type shown in Fig. 5(b) (see Section 4), grass occurred only when the tube temperature was below about 40°C. With permanent gases the appearance of grass was unaffected by changes in pressure and was also insensitive to changes in the magnetic field strength. No satisfactory theory of its occurrence has been evolved, but it is suggested that it may be associated with the motion of the cathode spot; alternatively its cause may be in the mechanism of formation of the arc spot.

Clean-Up of Gas

Considerable clean-up of gas took place during the discharge, severely limiting the life of any scaled-off tube which depended for its satisfactory operation on the presence of permanent gas. Some reduction of the clean-up was obtained by

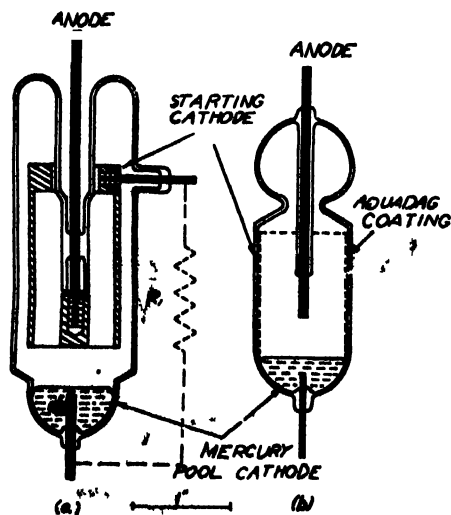


FIG. 5. Three-electrode tubes with anode, starting cathode, and mercury pool cathode; (a) with external resistance, (b) with internal resistance.

preliminary saturation of the walls with gas and by the addition of a buffer volume, but the longest life so achieved was only 50 hours. The effects of clean-up were avoided in tubes described in Section 4, which would start running on mercury vapor at room temperature. Attempts to run tubes on iodine and water vapor were unsuccessful.

4. ARC DISCHARGES IN MERCURY POOL TUBES

By addition of a mercury pool to the coaxial cylinder electrodes described in the preceding section a tube was obtained which operated on mercury vapor at room temperature and in which the discharge always took the form of an arc. Such tubes are shown in Fig. 5. The mercury pool acted as the cathode taking the main current and the outer cylindrical electrode merely as a starting cathode. Between the starting cathode and the anode a glow discharge was initiated by the magnetic field; bombardment of the mercury by ions from this discharge caused an arc to strike on the mercury surface.

These tubes operated satisfactorily at ambient temperatures between a few degrees C and 80°C, but below about 30°C unsteadiness occurred in the current trace on first switching on, disappearing as the tube warmed up. To ensure that the arc regularly formed on the mercury and not on the starting cathode, a resistance was placed in the starting cathode circuit. This was either an external resistance as in Fig. 5(a), or an internal resistance as in Fig. 5(b) where the starting cathode was a thin coat of Aquadag or sputtered tungsten. The latter type of tube had a life of only a few hours under the test conditions, i.e., as in a radar modulator, failure resulting from one of two causes; sometimes the action of the arc spot as it moved around the line of contact of the mercury and the glass separated the coating from the mercury pool; sometimes sputtering from the anode during current reversal after passage of the main pulse so lowered the resistance of the starting cathode that the arc spot formed on it instead of the pool, the voltage drop in the tube being then much higher and arcs sometimes forming again between the magnetic field pulses.

Of all types tested, the tube of Fig. 5(a), with the external resistance between starting cathode

and pool, gave the most satisfactory operation. Its life was limited by the formation of a conducting path of sputtered material and mercury between the pool and the lead to the starting cathode, with consequent arcing on the starting cathode as described above. Such tubes have handled powers up to two megawatts from a power supply of 10–20 kv at repetition frequencies up to 1200 c/sec. with pulse lengths from 1 to 10 μ sec. One had a life of 150 hours; the conducting path mentioned above could then still be broken up by shaking. The interval between the beginning of the magnetic field pulse and the beginning of the current pulse through the tube, which was from 20 to 50 μ sec., never varied from pulse to pulse by more than 2 μ sec.; the variation decreased as the supply voltage was increased, its average value being less than 1 μ sec. with a supply voltage of 10 kv. A magnetic field pulse of peak intensity 350 oersted was sufficient to fire the tube.

The value of the resistance R_1 was not critical, values from 0.01 to 0.5 megohm being satisfactory. Operation was not affected by tilt of the tube so long as the mercury did not touch the starting cathode. Alignment of the electrodes was not critical, nor was that of the field coil.

Graphite, which does not amalgamate with mercury and sputters at a slow rate in mercury vapor, was found to be a suitable material for starting cathode and anode. The clearance between these two electrodes and the pool was made greater than $\frac{1}{4}$ ", since "stalactites" of mercury sometimes grew downwards from them towards the pool. For a similar reason, the clearance between the starting cathode and the glass was kept greater than $\frac{5}{16}$ ".

In operation of the tube, gases are cleaned up and it was found that rigorous outgassing or very low pressures were not necessary before sealing off.

Figure 6 shows the oscillograph traces of tube current, anode potential, and starting cathode potential; these may be interpreted as follows. When the magnetic field reaches the critical value a glow discharge is initiated between the anode and the starting cathode, and the potential of the latter rises. About 0.7 μ sec. later the arc strikes on the mercury, and the starting cathode potential drops nearly to the cathode

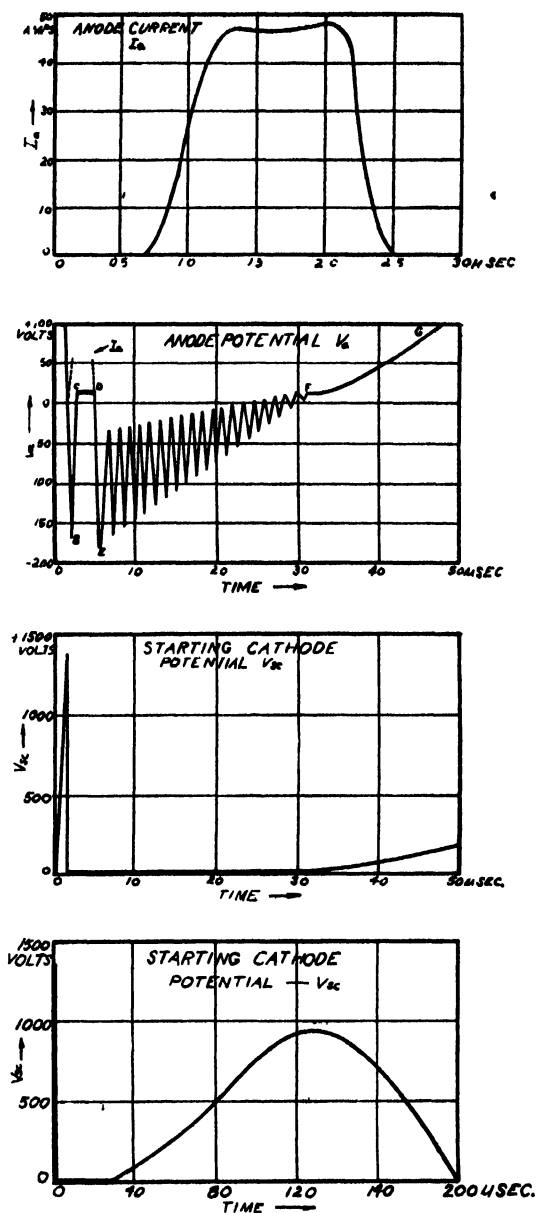


FIG. 6. Oscillograph traces of tube current, anode potential, and starting cathode potential.

potential (zero), remaining at this value while the discharge continues, and afterwards so long as the anode is negative. When the anode becomes positive the starting cathode potential also rises until deionization in the inter-electrode space is complete, when it falls again to zero.

It was observed that on first operating new tubes of this type the arc spot at first formed on

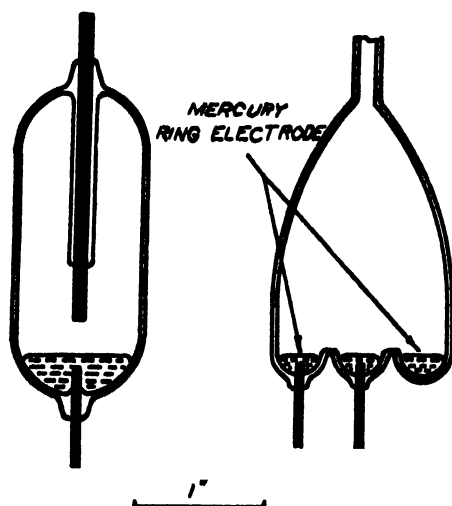


FIG. 7. Tubes with electrodes other than coaxial cylinders.

the free surface of the mercury pool, then formed only with difficulty until (after about 10 minutes) the mercury had begun to wet the glass; the spot then formed regularly at the meniscus.

Other Mercury Pool Tubes

During work leading up to the design of the tube just described, it was found that discharges could be initiated between electrodes whose geometry differed widely from the concentric cylindrical types and the configurations used by Penning.^{4,5} Some examples are shown in Fig. 7, all of which required a permanent gas in addition to mercury vapor for operation at room temperature and were thus subject to clean-up troubles.

5. INTERRUPTION OF CURRENTS BY REMOVAL OF THE MAGNETIC FIELD

The general features of the glow discharge initiated by a magnetic field have been discussed

in Section 3. It was found that, so long as the discharge did not pass into an arc, reduction of the magnetic field below the critical value interrupted the current through the tube, causing it to fall to a negligibly small value in less than one microsecond. Once the arc had been formed, however, removal of the magnetic field caused no change in the tube current, which continued to flow until the condensers were discharged.

Some attention was given to the production of a tube which would switch moderate or heavy currents off as well as on. It will be seen that the chief problem involved in this was to prevent the transition of the discharge from a glow to an arc. Concentric cylinder two-electrode tubes were used, and to prevent arc formation it was found necessary to keep the supply below about 3000 volts; this limit did not depend much on the current. Currents up to 50 amperes were interrupted in this way.

It is possible that in a tube with a large cathode area, containing a gas such as nitrogen, which has a large normal cathode current density, large currents could be switched on and off, passing with a few hundred volts drop across the tube. A cold cathode tube might thus be obtained which would do what is now only done with difficulty by high vacuum hot-cathode tubes.

We wish to acknowledge the extensive co-operation and help of Professor V. A. Bailey, the C.S.I.R. Radiophysics Laboratory, Sydney, and of Professor L. H. Martin, L. G. Parry, and W. R. Watson. While this work was being carried out, one of us (J.M.S.) was a holder of a Commonwealth Research Fellowship and another (K.R.M.) was employed by the Radiophysics Laboratory.

Depreciation of Fluorescent Lamps without Passage of Current

H. C. FROELICH

Lamp Development Laboratory, General Electric Company, Cleveland, Ohio.

(Received April 9, 1946)

Fluorescent lamps with thin and with partial coatings of phosphor on bulbs of ultraviolet transmitting glass were irradiated with Hg 2537 radiation. Comparison with similar lamps which were burned showed a decline in lumen output which was of the same general character in all experiments, except for different time factors. This indicates that the process of lamp depreciation is fundamentally of photochemical nature, involving excited mercury, but not requiring either Hg ions or neutral atoms. It is assumed that superficial oxygen ions of the phosphor crystal react with Hg* to form links of the type-SiO-O-Hg-Hg., thus tying mercury to the crystal with a resulting loss of phosphor brightness.

INTRODUCTION

IT is common knowledge that the light output of fluorescent lamps declines on burning. It has also been shown that this depreciation is caused by a screen of condensed mercury and mercury compounds in an extremely thin though discontinuous surface spread; it is not the result of a breakdown of the phosphor crystals in their bulk.¹

It seemed to be of fundamental interest to find out whether this depreciation is necessarily associated with the passage of current through a lamp, or whether it could also be brought about simply by a photochemical reaction involving excited mercury or perhaps excited phosphor. In the absence of mercury the phosphors do not depreciate. If a photochemical reaction should occur between mercury and the solids of glass wall or phosphor, then a new approach to the study and understanding of phosphor quality and "stability" (lamp maintenance) would have been found. Such reactions would have to be confined to the surface of the crystals because the bulk (or interior) of the crystals remains unchanged in brightness, and also because of the ease with which phosphors can be demercurized.

The problem, then, was to expose phosphors to the action of excited mercury and follow their brightness changes with time. Two types of devices were used for the purpose, the limitations of which with respect to rigorous treatment will be pointed out later. However, the accuracy desired for this exploratory work is believed to be sufficient to establish either the presence or ab-

sence of the effect; and the simplicity of the set-ups was a persuasive factor in deciding against more complicated apparatus.

EXPERIMENTAL

The first type of experiment has already been described recently.¹ In a jacketed lamp device the phosphor was irradiated from within, exposed to both $\lambda 2537$ and excited mercury vapor. The light output fell 25 percent during about 600 hours of irradiation, but the recovered phosphor, after proper demercurization, was found to be as bright as it was originally.*

* External to a lamp, the brightness of phosphor powder was measured as follows: A layer of powder about 2 mm thick, with flat and well-defined surface area, was irradiated with a beam of $\lambda 2537$ under oblique incidence. The amount of visible light generated by the excited phosphor was then measured with corrected photo-cells in terms of the brightness of an arbitrary standard of equal area, measured similarly and immediately before or after the sample. The accuracy of measurement was $\frac{1}{2}$ percent or better. A differential method might be more desirable.

Unfortunately, the mercury resonance line $\lambda 2537$ is the only monochromatic radiation which can be conveniently and abundantly produced in the spectral region where phosphors for use in F lamps are most sensitive. This fact necessitates complete absence of mercury and its compounds, solid, liquid, or gaseous, before the phosphor can be excited for correct measurement. With an equally good source of, say, $\lambda 2500$ or $\lambda 2600$ radiation, disturbance by mercury vapor would be eliminated but a reduction in brightness of non-demercurized phosphor would still be found. It would be in proportion to the area and density shaded by condensed Hg and Hg compounds, provided the latter are opaque to that radiation, which is likely to be the case. Demercurization is easily effected by heating the phosphor in a good vacuum to low temperatures (300°C or even as low as 150°C).

Removal of phosphor from a lamp involves complete random reorientation of the billions of individual crystals comprising the coating. Such bulk measurements, then, showed no significant difference in brightness between demercurized phosphors from unburned lamps and those from lamps several hundred or thousand hours old.

¹ H. C. Froelich, Trans. Electrochem. Soc. **87**, 365 (1945).

In the second type of experiment the same principle was used but the technique was reversed. The phosphors were irradiated from the outside, through ultraviolet transmitting glass. Phosphor brightness was measured in terms of

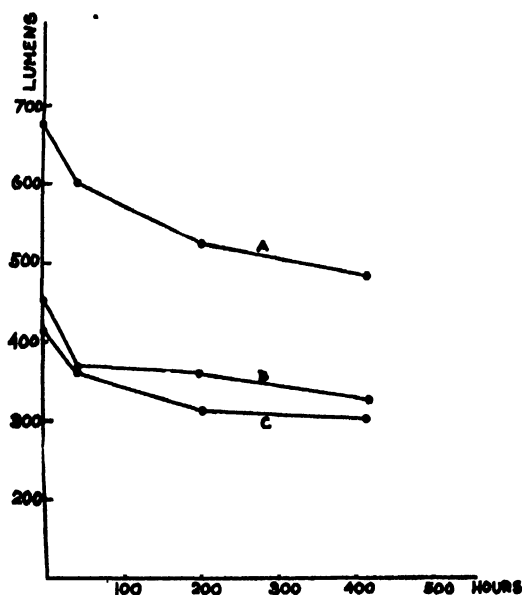


FIG. 1. Depreciation of half-coated lamps.

lamp brightness with current passing through the lamp, but without disturbing the orientation and the bulk of the phosphor crystals. Regular "white" phosphor was coated on 18" T8 germicidal bulbs of Corning No. 9741 glass, and regular lamps were made. The lamps were seasoned, photometered in a sphere for their lumen output, and then divided into several groups. One group was stored to be re-read at intervals to check accuracy. Another group was burned normally and also photometered at certain intervals. The last group was irradiated with $\lambda 2537$ in the following manner: Each lamp was surrounded by 6 regular germicidal lamps which were burned in order to provide a fairly uniform and reasonably high amount of $\lambda 2537$ radiation. No attempt was made to measure the wattage and density of $\lambda 2537$ radiation actually entering the test lamp. It was estimated, however, that it should be of the order of one-half to one-fourth of the amount generated in a normally burning lamp. The lamp to be irradiated was not burned except for the time when it was being measured.

In the first series the irradiated lamps were only half coated—i.e., half around their circumference but full length. In the second run, a uniform coating of less than normal thickness was applied all around the bulbs. A few controls of fully coated lamps in germicidal as well as lime glass were also made which were only burned. All lamps were photometered in a sphere at constant current of .300 amp. The data obtained are averaged and considerably condensed in Tables I and II.

Figures 1 and 2 show plots of the data in Tables I and II. It is apparent that the irradiated lamps decline in their output and follow the same qualitative pattern as the burned lamps. Because of the nature of the 9741 glass the depreciation curves of burned lamps cannot be expected to follow the same uniform pattern as regular lead or lime glass lamps. The scale of the time ordinate for the irradiated lamps in groups C had to be reduced in order to make the similarity apparent; this was pointed out above. The reduction factor in Fig. 1 was 3.6, while that for Fig. 2 was 2.6, as found by trial. To make the analogy more complete, the irradiated lamps were rejuvenated by application of external heat in a tube furnace at 300°C. The averaged lumen values were as follows:

After 2 hours	After 1501 hours	Rejuvenated
409	302	427

The apparent slight improvement of the rejuvenated lamp over the 2-hr. reading is believed to be caused, at least in part, by a slight deterioration during the first 2 hours.

Finally, a test was devised where depreciation and no depreciation could be produced in a single lamp. A 36" T8 lamp was made with a thin coating of white phosphor on a bulb of 9741

TABLE I. First run with half-coated lamps of normal coating thickness. (9741 glass). Averaged lumens at 0.300 amp.

Lamps	Depreciation due to burning for			
	2	44	204	421 hours
A. Full coat	675	600	523	480
B. Half-coat	452	367	359	323
C. Half-coat	Depreciation due to irradiation for			
	2	166	736	1501 hours
	409	361	311	302
Hours of irradiation, divided by factor 3.6		46	204	416

glass. One-half of the lamp was surrounded by 6 regular germicidal lamps, while the other half was surrounded by 6 lime glass lamps which generated light and heat but no external ultraviolet; these lamps were also of the 18" T8 (15-watt) size. All 12 lamps burned in order to produce nearly uniform radiation and temperature distribution within and around the test lamp in the center. Again, the test lamp did not burn except for the time when it was being measured after removal from the "light bath." It proved impractical to take sphere readings of lumen output. The two ends of the lamp, therefore, were photometered against a known standard by taking readings with a cell box arrangement which caught the light from about two-thirds of each section. After each interval of irradiation the test lamp was burned for at least 15 minutes before output readings were taken. The lamp was photometered at constant current of 0.350 amp. and under these conditions the meter deflections should be proportional to the lumen output. Initial readings for the two ends of the lamp were not equal because a coating thickness differential existed. It was decided to irradiate the thicker coated, brighter section of the lamp

TABLE II. Second run with full coated lamps of less than normal coating thickness. (9741 glass). Averaged lumens at 0.300 amp.

	2	168	305	360	928 hours
A. Normal coating, lamps burned	690	607		536	
B. Thin coating, lamps burned	550	458		420	
C. Thin coating, lamps irradiated	539		443		425
Hours of irradiation, divided by factor 2.6	.8		117		357

with $\lambda 2537$, because this might produce less favorable results (lower depreciation).

In addition to several other effects, a noticeable change of color was observed in the section which was irradiated with ultraviolet; the color temperature went up by a few hundred degrees K (estimated). The decline in output, however, was several times greater than might be expected on that basis alone. This and the other effects observed require further study.

Averages of several measurements are given in Table III and plotted in Fig. 3. Again it is apparent that the light output of the ultraviolet-irradiated section declined as though the lamp had been burned with normal passage of current,

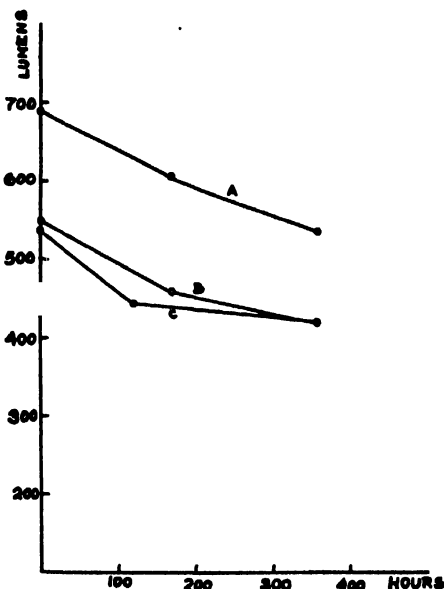


FIG. 2. Depreciation of thin coated lamps.

while the other half of the lamp remained substantially unchanged. The slight departure (up and down) from a straight horizontal line reflects the accuracy of measurement.

DISCUSSION

The experiments have given two results which are of fundamental interest. First, they have shown that fluorescent lamps can depreciate in the presence of excited mercury alone, without actual passage of current and in the absence of positive ions. Secondly, they have shown in a semiquantitative way that the depreciation due to excited mercury follows the same pattern as the depreciation of burning lamps: the decline is relatively greater during the initial periods than during the later stages. Excited mercury was produced by irradiation of its vapor with its resonance radiation $\lambda 2537$. This seems to prove that the phenomenon of F lamp depreciation is primarily of photochemical nature, involving excited mercury and/or excited phosphor. It would be of great interest to separate the two possi-

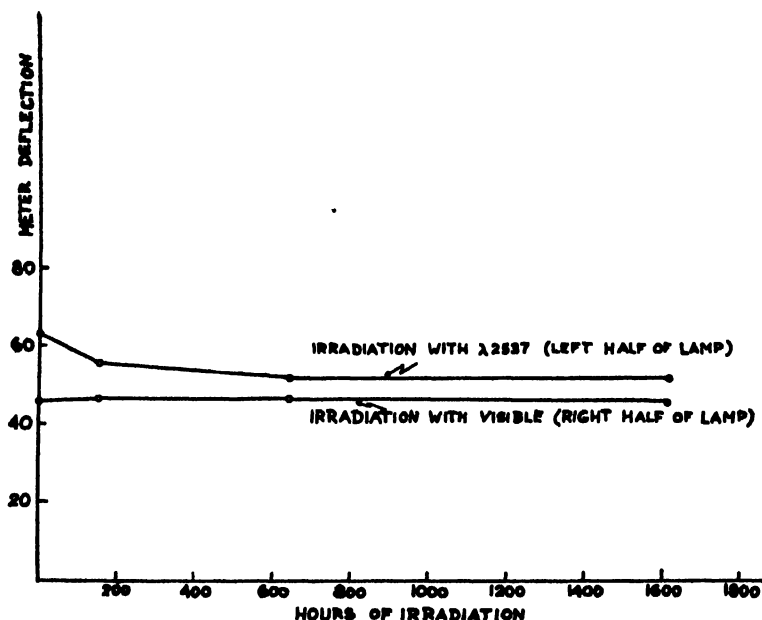


FIG. 3. Depreciation and stability of double-irradiated lamp.

bilities in repeat experiments with other than mercury resonance radiation. However, the latter should be near the peak of phosphor sensitivity, i.e., close to $\lambda 2537$.

One of the chief limitations of the present experiments lies in their semiquantitative nature. In plotting the curves with a linear time abscissa it is assumed that the wattage input of $\lambda 2537$ into the test lamp remains constant. Actually, the output of the germicidal lamps declines gradually (about 25 percent) during the course of the experiment, and the transmission to $\lambda 2537$ of the glass of the test lamp also declines through solarization. Therefore, the amount of $\lambda 2537$ actually getting into the test lamp becomes progressively smaller. In a quantitative experiment this amount should be determined, at least from time to time, and the output curves should be plotted against watt-hours of $\lambda 2537$ rather than against time alone. Such a plot will show a change of absolute slope compared with the present data, but qualitatively the decline pattern will remain unchanged. The scope of the present paper did not attempt to go beyond these limitations.

What are the forces which cause the deposition of mercury and its compounds on the phosphor, thus masking in part its otherwise unaffected capacity as a light transformer? One is forced to assume that a surface reaction (or, if one wishes,

"deterioration") does take place between the phosphor or glass and excited mercury atoms. This reaction does not proceed into the interior of the crystals, but is confined to the surface; it requires a relatively low energy of activation, and is reversible. If it were to go deeper, it would necessarily result in a physical destruction of the phosphor. Recovery of initial phosphor brightness indicates that such disintegration does not take place. Exact measurements of the extinction coefficient for $\lambda 2537$ in silicate, etc., phosphors would be another means of following the stability of these crystals. To the author's knowledge, such measurements have not been made with fine powders and $\lambda 2537$ radiation.

In the absence of mercury, phosphors do not depreciate; in its presence, they do it to varying degrees. There are phosphors which give good maintenance in a lamp, and others of about the same chemical composition which show bad maintenance. The tangible and intangible factors which play an additional role in the art of phosphor manufacture cannot be discussed here.

TABLE III. Averaged meter deflections at 0.350 amp.

	2	162	647	1625 hours
Ultraviolet-irradiated section	63	55	51	51
Visible-irradiated section	45.5	46.2	46	45

Suffice it to say that chemical composition alone is not a criterion for phosphor performance in a lamp. If a phosphor gives good maintenance, less mercury reacts with it and deposits on it compared with phosphors of poor maintenance. One might even speak of "mercurophobic" and "mercurophilic" phosphor surfaces although, strictly speaking, a lamp coated with phosphor of true mercurophobic surface should have no depreciation whatever.

It is easy to eliminate unexcited, atomic mercury as the agent responsible for deterioration. Atomic mercury could reach the phosphor surface through condensation from the vapor phase, as a wetting or non-wetting continuous liquid or in form of innumerable discontinuous droplets. Either assumption can be shown to be absurd. Consider, for example, a regular 40 watt, 48" T12 White lamp which has an inner (geometric) surface of about 1320 cm², contains about 2.2 g of phosphor and about 25 mg mercury. If the mercury were spread out into a uniform cylindrical film, it would form a layer of about 52 atoms or 156A thickness. It is inconceivable that the large surface tension forces would permit the free existence of a film stretched to that extent. Electron microscope pictures of metal films in thicknesses up to 100A or more show that they are completely broken up into separate particles.² Differences in thermal coefficients of expansion of metal and substrate likewise would cause rupture and coagulation of the "film." The concept of a continuous film of mercury seems untenable.

Simple condensation in form of innumerable tiny droplets also has its limits. If it is assumed that the normal 10 or 20 percent lumen loss of a fluorescent lamp be due merely to the shadow cast mechanically by subdivided mercury droplets, then the 25 mg of Hg in a 40-watt lamp should be expected to cover an area of 10 to 20 percent of the lamp surface or 132 to 264 cm². Assuming the droplets to be spheres, it can be calculated that this would require a subdivision into particles of the order of .1- to .05-micron diameter. Such small drops, however, would have a greatly increased vapor pressure compared with the larger drops from which they form. The

well-known Thomson equation permits calculation of these increases in vapor pressure.

$$\frac{dp}{p} = \frac{2\sigma M}{r \times \text{sp.g.} \times RT}$$

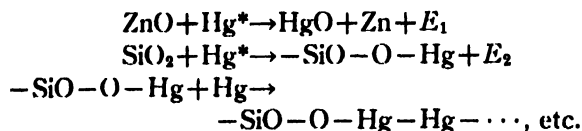
$$(\sigma_{\text{Hg}} = 450 \text{ dynes/cm}; M = 200; r = 5 \times 10^{-6} \text{ cm}; \text{sp.g.} = 13.55; R = 8.3 \times 10^7 \text{ erg/degree mole}; T = 323).$$

The equation holds so long as the surface tension σ_{Hg} is independent of the radius; in other words, r must be large against atomic dimensions. Substituting the proper numerical values it follows that a mercury particle of about .1-micron diameter (1000A or about 330 Hg atom diameters) would have a 10 percent higher vapor pressure at 50°C than a drop of 1-mm diameter. Such small particles would be thermodynamically unstable and would disappear by distillation. A few large particles would grow at the expense of most small particles. It would follow, then, that a lamp whose lumen output declined 10 or 20 percent should regain its initial brightness if it were simply left lying for weeks, months, or years; this has not been found by experiment. Conversely, a new lamp has likewise not been found to depreciate if it was neither burned nor irradiated.

Mercury droplets of thermodynamically stable size can account for not more than about 1 percent lumen loss if the latter is caused by mechanical shading of the phosphor against $\lambda 2537$. The calculation left entirely untouched such energetic difficulties as are connected with seed and phase formation when vapor atoms condense to form stable sub-micron particles. The situation becomes even more difficult if the (geometric) phosphor surface is considered instead of the glass surface. A 40-watt (G.E.) lamp contains about 2.2 g of white phosphor of sp.g. 4.5. Particles of all sizes from 1 micron and lower, up to 4, 5 microns or more are represented. Their total number is of the order of 10^{10} , and their total (geometric, not true) surface is of the order of 10^4 or 10^5 cm²—many times the area of the glass on which they rest. To summarize, the depreciation cannot possibly be produced by chemical action or by condensation from the vapor phase of neutral, atomic mercury.

² R. G. Picard and O. S. Duffendack, J. App. Phys. 14, 291 (1943).

This leaves excited mercury atoms as responsible for a reaction with the phosphor surface which can tie them up in a form in which their vapor pressure is lower than that of free Hg at the same temperature. Excited mercury could attack mercuriphilic phosphors or compounds by a reaction of the following types:



If we abandon the narrow chemical viewpoint of oxides and consider the phosphor or glass lattice as a whole, it is known that the latter consists of SiO_4 (or WO_4 , etc.) tetrahedra with interspersed Zn, Mn, etc. ions or atoms. The crystal surface layer contains more ions of oxygen than of any other species, and unsaturated oxygen valencies may conceivably become tied up with Hg^* . Existing stronger bonds may even be broken by the energy available from excited Hg (4.9 ev or 113 Cal.). Excess energy would be dissipated as heat as indicated by E in the equations. Thus the oxygen atoms may become a bridge linking Hg atoms to the bulk of the phosphor crystal by semi-chemical forces. Further growth involving $-\text{Hg}-\text{Hg}$ bonds is quite plausible. After one or more additional "layers" of Hg atoms, if such should be present, these semi-chemical forces could be assumed to deteriorate into physical adsorption, providing a condensed and opaque phase of sufficient density and stability to account for 10 to 20 percent absorption of $\lambda 2537$. Thus, only a fraction of the 25 mg of mercury in a 40-watt lamp would be required for the entire process of deterioration, and this is borne out by experience. Visual inspection of lamps of many hundreds of thousand hours burning indicates that a good portion, if not the bulk, of the mercury is still present in form of free droplets. The condensed phase of mercury would also provide an effective barrier for more excited atoms to react with the silicate surface underneath; the latter is then protected against further deterioration and the loss of brightness should occur at a greatly diminished rate, which is also observed. Complete clean-up of mercury by the phosphor occurs only when the

amount of mercury originally present was very small, or for other non-pertinent reasons.

It should be noted that this proposed mechanism of deterioration differs from somewhat similar suggestions which, however, go so far as to assume complete chemical reduction of the crystal surface to elemental zinc, silicon, etc. According to this theory, the oxygen is disposed of by assuming that it is liberated in gaseous form and is then free to diffuse and react elsewhere in the lamp. The loss of brightness is believed to be caused by the absorption of $\lambda 2537$ by this opaque coating of elemental zinc, silicon, etc. The present tests indicate that this assumption is erroneous. The recovery of initial lamp brightness after simple application of external heat of depreciated lamps speaks against the presence or formation of material with a high vapor or decomposition pressure. Moreover, silicon in its elemental form is a very stable substance which reacts with oxygen only at rather high temperatures. It is hard to visualize a reaction which would make it innocuous when the phosphor is heated in a vacuum to 150°C . On the contrary, the tests on the full recovery of phosphor and lamp brightness indicate that a permanent deposit could not have formed on the phosphor.

The change in color of the ultraviolet-irradiated white lamp, which was mentioned above, can be understood by assuming that the oxygen atoms or ions in tungstates are bound more tightly than in silicates. A relatively stronger attack by excited mercury atoms on the Zn-Besilicate phosphor compared with the magnesium tungstate component would result in a shift of color temperature toward higher values.

These speculations have been advanced chiefly in order to show what possibilities exist for chemical reactions which would stop at the crystal surface, tie the mercury to it, yet make it recoverable by such mild treatments as heating in a vacuum to low temperatures. With reference to the performance of fluorescent lamps, on the other hand, the term "depreciation" covers quite a complexity of processes. Each of these is contributory, though not individually responsible for the entire loss of lumen output. They include formation and deposition of a number of mercury compounds produced from excited mercury and

a variety of gaseous trace impurities left or formed in the lamp. This has been discussed in detail by Kenty and Cooper.³

SUMMARY

It has been shown that the process of lamp depreciation is fundamentally of photochemical nature, involving excited mercury atoms and the surface ions of phosphor or glass. Fluorescent

³ C. Kenty and J. Cooper, Trans. Electrochem. Soc. **87**, 287 (1945).

lamps containing mercury could be depreciated simply by irradiation with the mercury resonance radiation $\lambda 2537$, without passage of current. Semiquantitatively, this depreciation is of the same magnitude and of the same character as that of burned lamps; it is greater during initial periods than during later stages. Theories and suggestions for the tie-up of mercury are advanced. Superficial oxygen ions of the phosphor crystals are assumed to function as a bridge, linking mercury atoms to the bulk of the crystal.

The Interpretation of Harker Syntheses

M. J. BUERGER

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 6, 1946)

An elegant way of studying the geometrical meaning of a Harker synthesis is to transform it into a diagram here termed the implication diagram. This is done by transformation of polar coordinates, the characteristics of the transformation depending on the number of operations in the cyclical group of the axial symmetry for which the Harker synthesis is prepared. The implication diagram has the important property that it is a map of the location of atoms in the crystal structure plus additional locations (ambiguities) which would give rise to the same Harker synthesis, plus satellitic locations. Fortunately the satellitic locations can be identified as such. Satellitic peaks are caused by powers of rotation operations and by reflections. Satellitic peaks occur in specialized locations which are discussed for the several cases. On the other hand, it is not generally possible to decide between ambiguities. An implication diagram which is a map of the crystal structure, with ambiguities, but without satellitic locations, exists for the following 18 of the 21 possible parallel-axial symmetries: $P2$, $P2_1$, $C2$, $H3$, $H3_1$, $H3_2$, $R3$, $P\bar{4}$, $I\bar{4}$, $P4_1$, $P4_2$, $P4_3$, $I4_1$, $H6_1$, $H6_2$, $H6_3$, and $H6_4$. This is also true of certain axial symmetries combined with reflections, such as $2c$ and $3c$. Only for the three axial symmetries $P4$, $I4$, and $H6$ do no implications exist without

satellitic peaks. The implication diagrams of the Harker syntheses $P(xy\frac{1}{2})$ for the space groups $H6_1$, $H6_2$, $H6_3$, and $H6_4$ and for Harker synthesis $P(xy0)$ for $R\bar{3}$ lack both ambiguities and satellitic peaks. For these space groups the implication diagrams are true maps of the crystal structure. Harker syntheses greatly exaggerate the electron densities of atoms in the structure, and for this reason their greatest usefulness lies in the location of the more compact atoms. Diffuse atoms, and therefore especially anions, provide Harker peaks of such low magnitude as to be often lost in the background. Background due to non-Harker interactions is discussed. An important characteristic of Harker syntheses, which is independent of any interpretation of their meaning with regard to the location of atoms in the crystal structure, is that they provide criteria for distinguishing space groups which cannot be distinguished in ordinary qualitative x-ray crystallography. This is because the Harker synthesis provides definite criteria for identifying reflection planes and rotation axes. It fails only to distinguish pairs of space groups which differ by a group of inversions alone. Harker synthesis thus realizes the ultimate possibilities of x-ray crystallography because it is quantitative.

INTRODUCTION

THE major difficulty besetting crystal structure analysis is the loss of phase relations in the recording of the diffraction. If these phases could be determined, all crystal structures could be solved. Patterson^{1,2} investigated the alterna-

tive question, namely, what information can be ascertained about a crystal structure if the phases of the diffraction spectra are unknown. He discovered the important function,

$$P(xyz) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hkl}|^2 \times \cos 2\pi(hx + ky + lz), \quad (1)$$

¹ A. L. Patterson, "A Fourier series method for the determination of interatomic distances in crystals," Phys. Rev. **46**, 372-376 (1934).

² A. L. Patterson, "A direct method for the determina-

tion of the components of interatomic distances in crystals," Zeits. f. Krist. **A90**, 517-542 (1935).

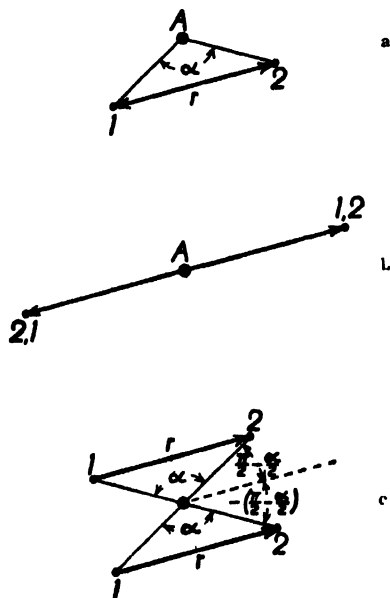


FIG. 1. Transformation of position of two points related by operation $A(\alpha)$ from crystal to Harker diagram and return. a: Relations in crystal. b: Relations in Harker diagram. c: Relations in *implication* diagram

and proved that this function is equivalent to another giving the average value of the product of the electron density at the two ends of a vector, as the vector, fixed in direction and length, is allowed to range over the crystal cell. This function has a high value when two atoms in the crystal structure are separated by this vector. Consequently (1) has maxima at points whose locations, with respect to the origin, define vectors which correspond to interatomic vectors within the crystal.

Harker² pointed out that, for crystals having any symmetry elements other than inversion centers, certain special regions of Patterson's three-dimensional function are the loci of the ends of vectors relating symmetrical equivalent atoms. These are therefore very useful in locating the positions of such atoms. For the purpose of locating atoms related by axes of symmetry, certain plane sections of the graphical representation of Patterson's function are useful, while certain lines are useful in aiding in the location of atoms related by symmetry planes.

² David Harker, "The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag_3AsS_3 , and pyrrargyrite, Ag_3SbS_3 ," J. Chem. Phys. 6, 381-384 (1936).

In this paper, the discussion is concerned with atoms related by axial symmetry elements, with or without symmetry planes parallel with the axes. For axial symmetry, Harker recommends the particular Patterson functions $P(xy0)$ and $P(xy\frac{1}{2})$. For these he gives the following form (axis assumed parallel to Z).

Screw Component of Axis = 0

$$P(xy0) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C \cos 2\pi(hx + ky), \quad (2)$$

where

$$C = \sum_{l=-\infty}^{\infty} |F_{hkl}|^2.$$

Screw Component of Axis = $c/2$

$$P(xy\frac{1}{2}) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C \cos 2\pi(hx + ky) \quad (3)$$

where

$$C = \sum_{l=-\infty}^{\infty} (-1)^l |F_{hkl}|^2.$$

It will be evident from what follows that these two cases, $P(xy0)$ and $P(xy\frac{1}{2})$, are not sufficient, and are not even recommended in some cases where they apply. A more general form for $P[xy(1/w)]$ is

$$P\left(xy\frac{1}{w}\right) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_1 \cos 2\pi(hx + ky) - \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_2 \sin 2\pi(hx + ky), \quad (4)$$

where

$$C_1 = \sum_{l=-\infty}^{\infty} \cos \frac{2\pi}{w} l |F_{hkl}|^2,$$

and

$$C_2 = \sum_{l=-\infty}^{\infty} \sin \frac{2\pi}{w} l |F_{hkl}|^2.$$

THE IMPLICATION DIAGRAM

The interpretation of Harker diagrams for certain axial symmetries is greatly simplified by a transformation of the data to another diagram which will be termed the *implication diagram*. The transformation is as follows: In Fig. 1a,

suppose that the two atoms, 1 and 2, are related by an operation $A(\alpha)$, of a group of rotations. (The notation here is that ordinarily used in discussions of crystal symmetry, namely, the axis of the rotation is designated by A and the amount of the rotation by α .) The atoms 1 and 2 determine two interatomic vectors, namely 1,2 and 2,1. Consequently the Harker diagram, Fig. 1b, contains two peaks, 1,2 and 2,1, which correspond to these vectors. Thus, to each distribution of atoms in the crystal there corresponds a unique distribution of peaks on the Harker diagram.

Consider the reverse transformation. What inference can be drawn from the Harker diagram regarding the distribution of atoms in the crystal? To find the possible atomic positions which have given rise to the Harker peaks, all that is necessary is to find two points in the crystal equidistant from axis A which subtend an angle α at A , and which are separated by distance 1,2= r . There are two solutions to this implication, Fig. 1c, unless $\alpha=\pi$, in which case the two solutions coincide.

The mechanism for transformation of any peak from the Harker diagram to the implication diagram is easily formulated. If the polar coordinates of the Harker point are r and φ (Fig. 2a), the polar coordinates of the corresponding two implication points (Fig. 2b) are

$$\left. \begin{aligned} s &= \frac{r}{2 \sin \frac{\alpha}{2}} = \frac{r}{2 \sin \frac{\pi}{n}} \\ \text{where } n &\text{ is defined as } \frac{2\pi}{\alpha} \\ \psi_1 &= \varphi + \left(\frac{\pi - \alpha}{2} \right), \\ \psi_2 &= \varphi - \left(\frac{\pi - \alpha}{2} \right). \end{aligned} \right\} (5)$$

An especially easy way of making use of this relation for certain axial symmetries will be discussed after certain other properties of the implication diagram are mentioned.

From (5) it is evident that each pair of points

on the implication diagram corresponding to one point on the Harker diagram is separated by an angle of $(\pi/2 - \alpha/2) + (\pi/2 - \alpha/2) = \pi - \alpha$ at the axis A . The following discussion will show that if each of the n points of a symmetrical set on a Harker diagram is transformed to the implication diagram, the points of the implication diagram with azimuth ψ_2 in (5) may be disregarded.

Harker syntheses are cosine summations, so that for each point, p_1 , at coordinates x_1, y_1 there is another point, p_2 , at coordinates $\bar{x}_1, \bar{y}_1$. The point, p_2 , subtends an angle π at the axis A with point p_1 . Because of the symmetry of the axis A , there is a third point, p_3 , which subtends an angle $\pi + \alpha$ with p_1 . Now consider the implication points corresponding with p_1 and p_3 . The second implication point of p_1 has an azimuth of $\varphi - \pi/2 + \alpha/2$; the first implication point of p_3 has an azimuth of $\varphi + \pi + \alpha + (\pi/2 - \alpha/2) = \varphi + 3\pi/2 + \alpha/2 = \varphi - \pi/2 + \alpha/2$. Thus the ψ_2 of p_1 is identical with the ψ_1 of p_3 . As a consequence, if all points of the axially symmetrical set of n Harker points are to be transformed into their respective implication points, it is sufficient to locate on the implication diagram merely one point with azimuth ψ_1 for each Harker point, since the second point with azimuth ψ_2 will appear as the ψ_1 point corresponding with another point of the symmetrical Harker set.

It follows from this discussion that to every Harker set of n axially equivalent points there corresponds a symmetrical set of n points on the implication diagram. Furthermore, these n implication points, except for the ambiguities discussed in the following sections, are the positions of atoms in the projection of the crystal structure on the plane normal to the axis.

The several kinds of ambiguities involved in implication diagrams will now be considered.

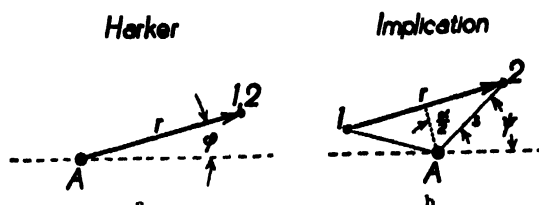


FIG. 2. Relations between polar coordinates of points on Harker diagram and implication diagram.

AMBIGUITIES IN THE IMPLICATION

Ambiguities Due to Centrosymmetry

Since, for every interatomic vector there exists the reverse vector, all Harker diagrams as

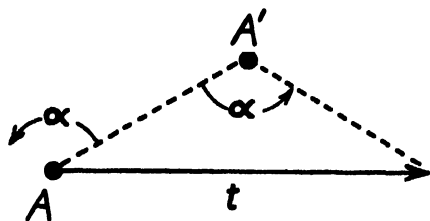


FIG. 3. The combination of an operation of rotation, $A(\alpha)$, with an operation of translation, t , and the consequent operation $A'(\alpha)$.

well as all implication diagrams are centrosymmetrical. Thus the group of symmetry operations of the axis A of the Harker diagram or implication diagram must contain the operation $A(\pi)$ whether the symmetry of the crystal axis A contains this operation or not. The only crystallographic case where it does not is for $n=3$ (not including the symmetry designated by "3"). This requires sixfold axes to appear on the Harker diagrams and implication diagrams corresponding to $n=3$ in the crystal. Thus for $n=3$ only, a twofold ambiguity arises in the interpretation of implication diagrams due to this cause.

Ambiguities Due to Translations

The interpretation of Harker diagrams and implication diagrams is complicated by the translation symmetry of crystals. This is fundamentally caused by the well-known fact⁴ (Fig. 3) that if the operation of a rotation, $A(\alpha)$, is combined with a translation t at right angles to the axis, the combination is equivalent to another rotation, $A'(\alpha)$, of identical magnitude about a parallel axis, A' . This causes the parallel symmetry axes (as well as the axes representing their subgroups of operations) of space groups to occur in several sets which are not necessarily translation-equivalent. These two axes are translation equivalent, however, for $n=6$, where they are related by a translation at right angles to

the axis; they may also be equivalent by a translation which is not perpendicular to the axis for other values of n when they occur in the centered and rhombohedral lattices.

The consequences of this for the interpretation of Harker diagrams and implication diagrams may be illustrated for the simple case of a one-dimensional "space group" represented by a linear array of twofold axes, Fig. 4b. The twofold axes occur in sets A and A' , which are not translation-equivalent. If an atom is placed at 1, then the twofold axis requires another at 2, and the translation reproduces these at $1t$, $2t$, etc. In addition to the interatomic vector $1,2$ and its reverse, there arises vector $1t, 2$, and its reverse, due to translation. The Harker diagram, Fig. 4c, then displays peaks at the ends of these two vectors, both of which radiate from the same origin. This kind of multiplication of peaks does not give rise to any ambiguity, for in transforming the Harker diagram to its implication, one need only consider any section of total length t (most conveniently the section within the range $\pm t/2$ of the origin). Only vectors corresponding to points within this range are distinctive, and

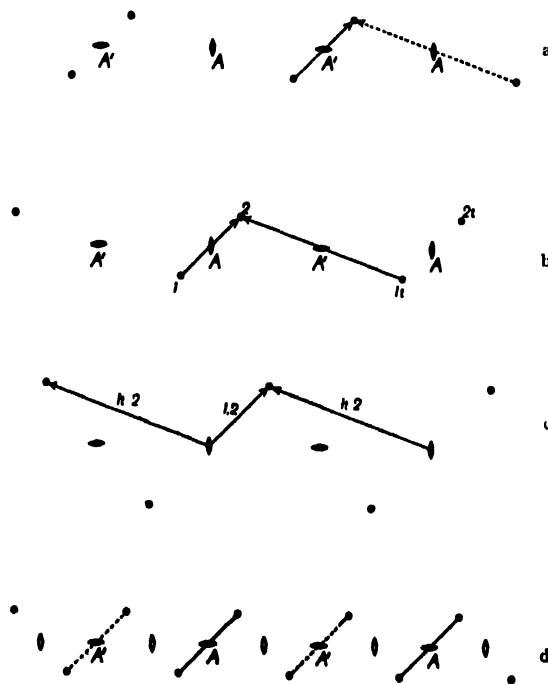


FIG. 4. Relations between crystal, Harker diagram, and implication diagram, and the cause of ambiguity. Locations of symmetry elements on implication diagram, which are not present in crystal, are indicated by outlined symbols.

⁴ Harold Hilton, *Mathematical Crystallography and the Theory of Groups of Movements* (Clarendon Press, Oxford, 1903), p. 147.

one therefore disregards vectors corresponding to points beyond this range.

There is, however, a second consequence of principle illustrated in Fig. 3. If the same distribution of atoms with respect to axis A , shown in Fig. 4b, occurs about axis A' , as shown in Fig. 4a, the same Harker distribution as that shown in Fig. 4c results. It is therefore impossible to determine whether the Harker diagram implies Fig. 4a or Fig. 4b. This is a general feature of implication diagrams, and it arises because the space group, in consequence of the principle illustrated in Fig. 3, contains several sets of non-equivalent axes of equal n . In space groups, the number of sets of such parallel axes depends on the symmetry of the axis. For uncentered lattices there are always four sets of twofold axes, three sets of threefold axes, two sets of fourfold axes, and only one set of sixfold axes. This number is reduced in centered and rhombohedral lattices by the factors 2 and 3, respectively. Let this number of sets be represented by m . Then, in transferring a Harker synthesis to the implication diagram, an m -fold ambiguity is involved, since one cannot tell which of the m sets of axes in the crystal relates the atoms so as to give rise to the Harker points.

In order to take care of this ambiguity, the implication diagram requires that the inferred locations of atoms shown in Fig. 1c be plotted about each of the m non-equivalent axes of the space group, as illustrated in Fig. 4d for this

TABLE II

	n			
	2	3	4	6
s	$\frac{1}{2}r$	$\frac{1}{\sqrt{3}}r$	$\frac{\sqrt{2}}{2}r$	r
ψ	φ	$\varphi \pm 30^\circ$	$\varphi \pm 45^\circ$	$\varphi \pm 60^\circ$
Shrinkage factor	$\frac{1}{2}$	$\frac{1}{\sqrt{3}}$	$\frac{\sqrt{2}}{2}$	1
Rotation	0°	30°	45°	$60^\circ \approx 0^\circ$

simplified case. This has interesting consequences. Since identical implications are plotted about non-equivalent axes, the latter take on the characteristics of translation equivalence in the implication diagram. This diagram therefore always has shorter translation periods than the crystal (except in the case of $m=1$).

Table I succinctly summarizes some important properties of the Harker diagrams and implication diagrams for the several axial symmetries. Unfortunately, for $n>3$, one cannot deal with sets of parallel axes of equal n . This is because the powers of the operations of such axes form subgroups of lower n . In the last column of Table I a superscript, r , appended to the symbol of a symmetry axis, designates a subgroup composed of the r th, $2r$ th, $3r$ th, etc., powers of the fundamental operation of the axis. All subgroups on the same line are identical in character with the group indicated by the symbol in column 2. The arrangement is intended to bring out the fact that for purposes of constructing and interpreting Harker and implication diagrams, atoms related by a 6_1 axis, for example, can be studied by the Harker synthesis $P(xy\frac{1}{6})$; or by the synthesis $P(xy\frac{1}{3})$, thus studying, in implication, atoms related by the second power of the operation of the screw; or by the synthesis $P(xy\frac{1}{2})$, thus studying, in implication, atoms related by the third power of the operation of the screw. Harker's original paper recommends the last two cases. The study should always be made by $P(xy\frac{1}{6})$, however, for the ambiguity factors are very different, namely 1, 3, and 4 for the powers 1, 2, and 3, respectively. When the implication of $P(xy\frac{1}{6})$ is prepared, there is no ambiguity due to translation, and the implication diagram is a map of the crystal structure projected on the plane normal to the sixfold axis.

TABLE I.

n	Axial symmetry	Harker synthesis	Primitive lattice	Ambiguity of location, m Centered or rhombohedral lattice	Subgroups of same n
2	2	$P(xy0)$	4	2	$4^1, 4^2, 6^1, 6^2$
	2_1	$P(xy\frac{1}{2})$	4	2	$4_1^1, 4_1^2, 6_1^1, 6_1^2$
3	$3, \bar{3}(=3+\bar{2}),$ $\bar{6}(=3/m)$	$P(xy0)$	3	1	6^1
	$3_1, 3_2$	$P(xy\frac{1}{3})$	3	2	$6_1^1, 6_1^2, 6_2^1, 6_2^2$
4	$4, \bar{4}$	$P(xy0)$	2	1	
	$4_1, 4_2$	$P(xy\frac{1}{4})$	2	2	
	4_2	$P(xy\frac{1}{2})$	2	2	
6	6	$P(xy0)$	1	—	
	$6_1, 6_2$	$P(xy\frac{1}{6})$	1	—	
	$6_2, 6_4$	$P(xy\frac{1}{3})$	1	—	
	6_3	$P(xy\frac{1}{2})$	1	—	

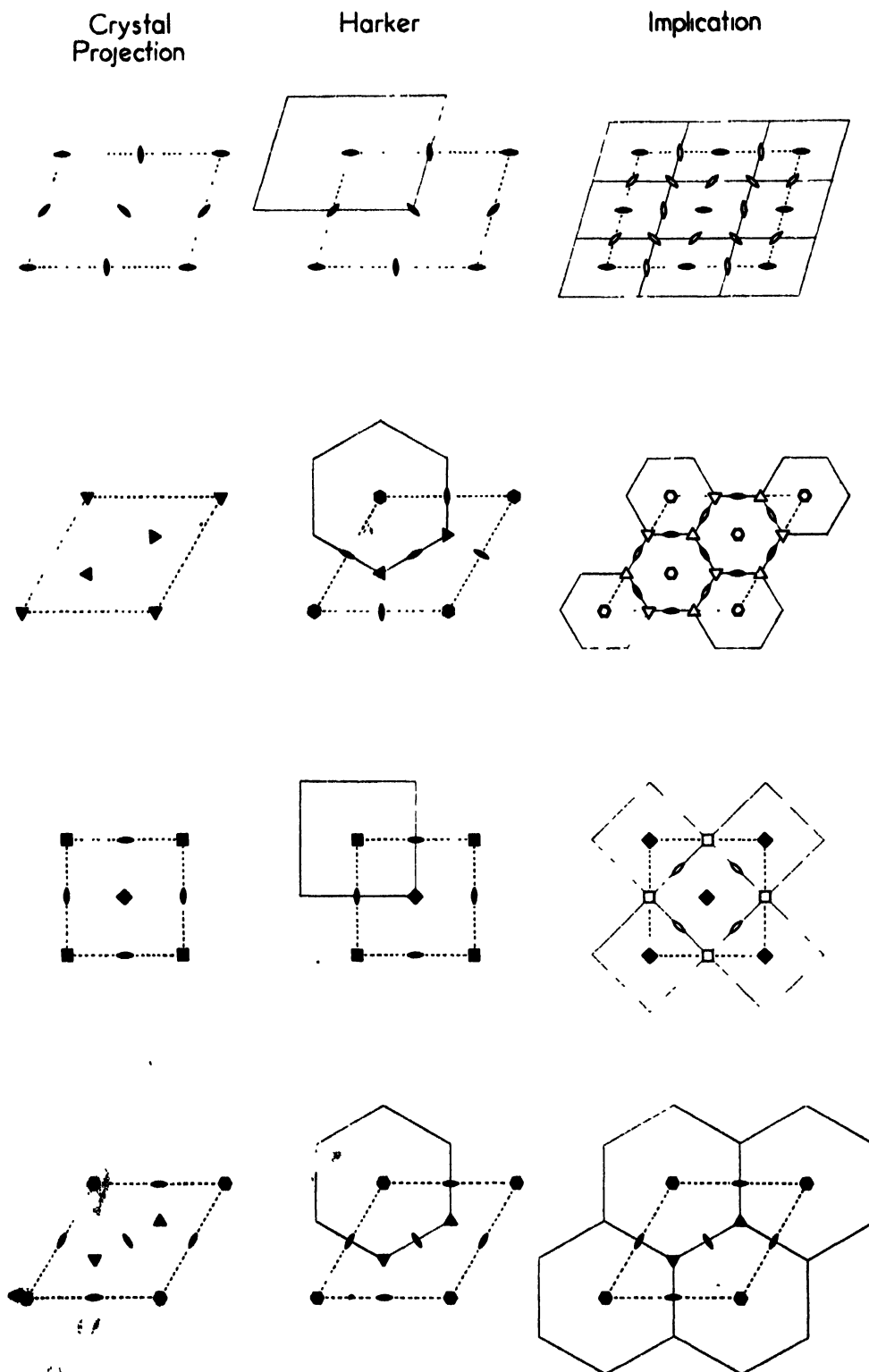


FIG. 5. Relations between the symmetry elements of the crystal, Harker diagram, and *implication* diagram for the four possible values of n . The outlines of the crystal cells are indicated by broken lines. The figure shows the domains (outlines in full lines) of the Harker synthesis and their transfer to the implication diagram for each case. The locations of new symmetry in the implication diagram for the case of primitive cells are indicated by outlined symbols.

CONSTRUCTION OF IMPLICATION DIAGRAMS

It is evident that an implication diagram can be constructed by taking the desired features, such as peaks, of the Harker diagram and transforming these, point by point, to an implication diagram. A better way of handling the manipulation, however, will be evident from the following discussion.

It will be observed from (5) that every point on the Harker diagram transforms into a point on the implication diagram by a shrinkage of its radial polar coordinate and a rotation of its azimuth. The shrinkage and rotation are constant for a given value of the axial symmetry parameter, n . The specific values of shrinkage and rotation for various values of n are given in Table II.

Evidently an implication diagram can be plotted from the Harker synthesis by transformation of the polar coordinates of the synthesis. A practical way of performing this is to plot the Harker synthesis to a scale reduced by the correct radial factor of Table II, rotate the plot by the correct angle given in Table II, and lay down the resulting plot on the m locations of similar axes discussed under "Ambiguities Due to Translation." The implication diagram constructed in this way shows all the detail of the Harker synthesis, instead of just the peaks.

No interpretation need take place up to this stage. Any proposed crystal structure can then be plotted directly on the features of the implication diagram. This procedure has the advantage of deferring any decisions about the meaning of doubtful features of the Harker diagram until they can be tested directly in comparison with the features of the proposed crystal structure.

Figure 5 shows the relations between symmetry elements in the crystal, the Harker diagram, and the implication diagram for the four possible values of n for pure axial symmetry. This figure also shows domains in the Harker synthesis and their transfer to the implication diagram for the several axial symmetries. Symmetry elements in the implication diagram shown in outline represent locations of symmetry not present in the crystal. Cell outlines are shown in broken lines.

SATELLITIC PEAKS ON IMPLICATION DIAGRAMS

Satellitic Peaks Due to Powers of Rotation Operations

The groups of operations represented by axial symmetries (pure rotations, screws, and roto-reflections) are cyclical. Hence only powers of the fundamental operation are members of the group. This gives rise to important consequences. In the case of screw symmetry, each atom can only be related by any selected operation of the screw to two other atoms, namely, one atom above and one atom below. There are thus only two Harker vectors from the atom due to this operation and its reverse. The two vectors have equal length and diverge by an angle of $\pi - \alpha = \pi - 2\pi/n$. For the case of $n=2$, the two vectors coincide, but they are distinct for all other cases. Thus, unless further complications occur, implication diagrams for screws are comparatively simple, and indeed, each peak on the implication diagram is the location of an atom, except for ambiguities.

No further complication occurs for symmetries with $n=2$ or 3, nor for the screws 4_1 , 4_2 , 4_3 , 6_1 , and 6_5 . Unfortunately, however, complications do arise for the axial symmetries 4 , 6 , 6_2 , 6_3 , and 6_4 . This is due to the fact that for these five symmetries (Fig. 6) there occur in the same plane not only the atom related to the original atom by the first power of the symmetry operation, but also one or more additional atoms related to the original atom by higher powers of the same symmetry operation (possibly reduced by a translation parallel to the axis). This gives rise to Harker vectors in addition to the two mentioned in the last paragraph. The several cases are illustrated in Fig. 6.

The additional vectors complicate the interpretation of Harker syntheses. The implication diagrams not only contain peaks at location of atoms, plus their ambiguities, but also additional peaks where there are no atoms. Of course, implication diagrams for such symmetries are easily interpreted if the crystal contains only one set of equivalent atoms, an extremely rare case indeed. In the more usual case of a crystal containing several sets of atoms, the diagram is more difficult to interpret. The interpretation is aided by taking account of the relations between

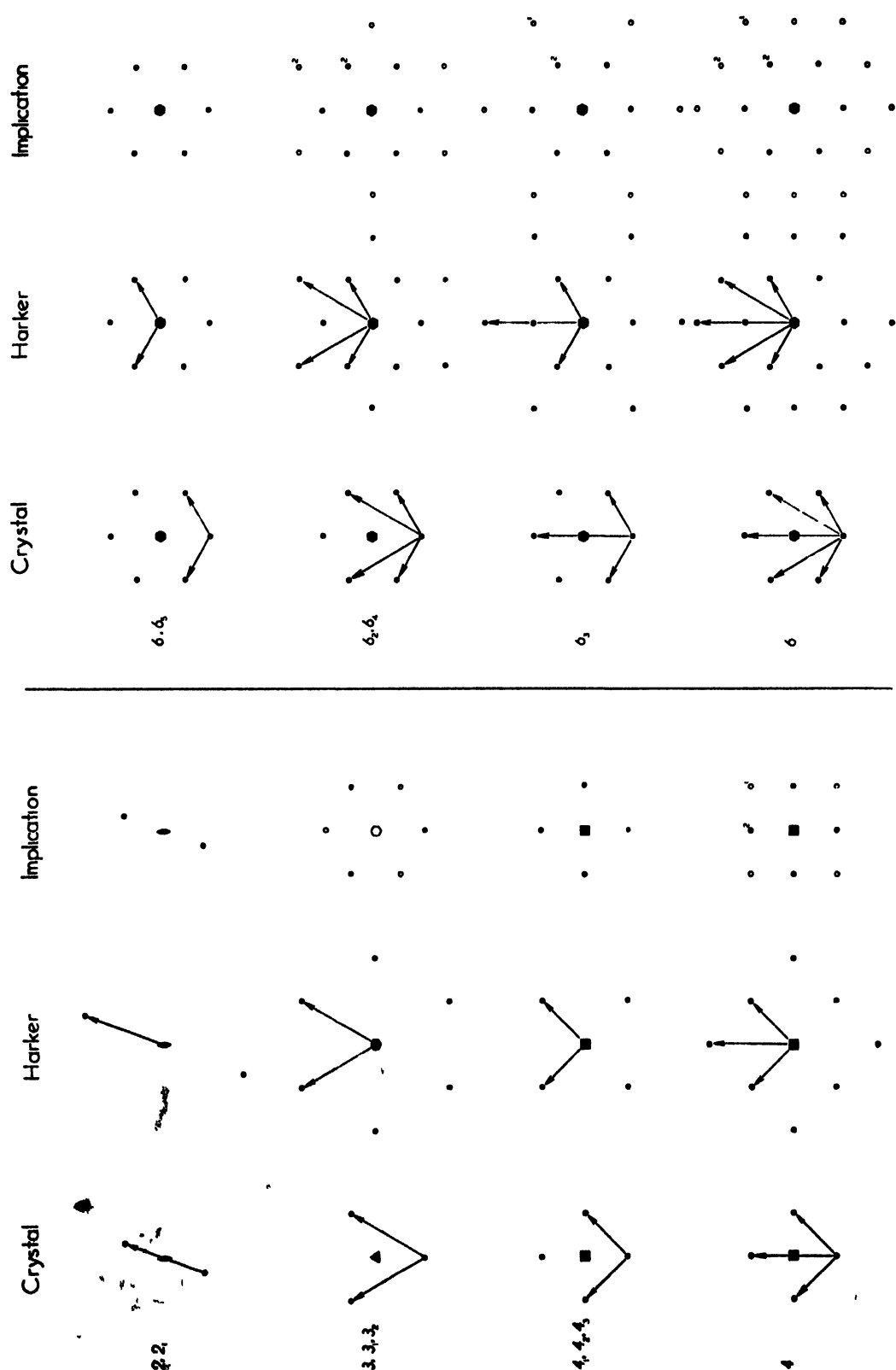


FIG. 6. Derivation of the point locations in the implication diagram for the several kinds of axial symmetry. Points and symmetry in the implication diagram which are not present in the crystal are indicated by outlined symbols. The relative heights of the several peaks, where these are different, are indicated by small numerals on the implication diagram.

the multiple sets of points, both with regard to geometry and comparative intensities.

The confusion due to the multiple sets of points is mitigated, for the symmetries 6_2 , 6_3 , and 6_4 , by the fact that Harker syntheses can be prepared for these symmetries for powers other than the first power. Thus, the symmetry 6_3 can be treated as 6_3^2 by using $P(xy0)$. The implication diagram then has the characteristics of $n=3$. Similarly, the symmetries 6_2 and 6_4 can be treated as 6_2^3 and 6_4^3 by using $P(xy0)$. The implication diagrams then have the characteristics of $n=2$. Each of these treatments involves the ambiguities characteristic of its n . It is possible to make use of this change in character of the ambiguity so as to provide a criterion for discarding satellitic peaks. This can be done by preparing implication diagrams for the several different powers of the same operation. Each implication contains the true implication plus a false implication due to ambiguities plus satellitic peaks. The peaks which these diagrams have in common are the true locations of the atoms, the false ones being, in general, different on the implication diagrams prepared for two different powers of the operation.

The characteristics of the possible implication diagrams can be formulated by considering the implication diagrams of the possible axial space groups instead of merely the individual axes. There are 21 space groups involving only operations about sets of axes parallel to a single line. The characteristics of the implication diagrams of these groups are set forth in Table III. The first column of this table lists the Hermann-Mauguin symbol of each of the 21 parallel-axial space groups. The second column provides the possible Harker sections for each such space group. The third column gives the possible values of n for which the implication can be constructed. For each of these values the axial symmetries and their ambiguity factor are listed in the next two columns. In the last column are shown the subgroups (if any) of the space group which give satellitic peaks in the implication diagram for that implication construction.

The table brings out the point that for several parallel axial space groups there exist implication diagrams which have neither translation ambiguity (i.e., $n=1$) nor satellitic peaks (no

interfering subgroups). These are $R3$, $H6_1$, and $H6_2$. The perfection of $R3$ in this respect is marred by the fact that there is an ambiguity due to centrosymmetry. The group $R\bar{3}$, however, is perfect in this respect. Evidently all axial space groups having $R\bar{3}$, $H6_1$, or $H6_2$ as subgroups have implication diagrams whose peaks are located at the positions of the atoms in the crystal structure. Thus implication diagrams exist for the following space groups which are maps of the crystal structure: $R\bar{3}$, $H6_1$, $H6_2$, $H6_3$, $H6_4$. The following axial space groups have trigonal implication diagrams which are maps of the crystal structure plus a centrosymmetrical

TABLE III.

Axial groups	Harker syntheses	n	axis	m	Subgroups causing satellitic peaks
$P2$	$P(xy0)$	2	2	4	
$P2_1$	$P(xy\frac{1}{2})$	2	2_1	4	
$C2$	$P(xy0)$	2	2	2	
	$P(xy\frac{1}{2})$	2	2_1	2	
$H3$	$P(xy0)$	3	3	3	
$H3_1$	$P(xy\frac{1}{3})$	3	3_1	3	
$H3_2$	$P(xy\frac{2}{3})$	3	3_2	3	
$R3$	$P(xy0)$	3	3	1	
	$P(xy\frac{1}{3})$	3	3_1	2	
$P\bar{4}$	$P(xy0)$	2	$4^2=2$	4	
$I\bar{4}$	$P(xy0)$	2	4^2	2	
	$P(xy\frac{1}{2})$	2	2_1	2	
$P4$	$P(xy0)$	$\begin{cases} 4 \\ 2 \end{cases}$	$\begin{cases} 4 \\ 2 \end{cases}$	$\begin{cases} 2 \\ 4 \end{cases}$	$\begin{matrix} P2 \\ P4 \end{matrix}$
$P4_1$	$P(xy\frac{1}{4})$	4	4_1	2	
	$P(xy\frac{3}{4})$	2	$4_1^3=2_1$	4	
$P4_2$	$P(xy\frac{1}{2})$	4	4_1	2	
	$P(xy\frac{3}{4})$	2	$4_2^3=2_1$	4	
$P4_2$	$P(xy\frac{1}{2})$	4	4_2	2	
	$P(xy0)$	2	$4_2^3=2$	4	
$I4$	$P(xy0)$	4	4	1	$I2$
	$P(xy\frac{1}{2})$	$\begin{cases} 2 \\ 4 \end{cases}$	$\begin{cases} 4^2=2 \\ 4_2 \end{cases}$	$\begin{cases} 2 \\ 1 \end{cases}$	$\begin{matrix} I4 \\ I2 \end{matrix}$
	$P(xy\frac{1}{4})$	4	4_1	2	$I4$
$I4_1$	$P(xy\frac{1}{4})$	2	$4_1^3=2_1$	2	
	$P(xy\frac{3}{4})$	2	2	2	
	$P(xy0)$	2	2	2	
$H6$	$P(xy0)$	$\begin{cases} 6 \\ 3 \\ 2 \end{cases}$	$\begin{cases} 6 \\ 3 \\ 2 \end{cases}$	$\begin{cases} 1 \\ 3 \\ 4 \end{cases}$	$\begin{matrix} H3, P2 \\ P2, H6 \\ H6, H3 \end{matrix}$
$H6_1$	$P(xy\frac{1}{6})$	6	6_1	1	
	$P(xy\frac{5}{6})$	3	$6_1^5=3_1$	3	
	$P(xy\frac{1}{3})$	2	$6_1^3=2_1$	4	
$H6_2$	$P(xy\frac{1}{3})$	6	6_2	1	
	$P(xy\frac{2}{3})$	3	$6_2^3=3_2$	3	
	$P(xy\frac{4}{3})$	2	$6_2^3=2_1$	4	
$H6_3$	$P(xy\frac{1}{6})$	$\begin{cases} 6 \\ 3 \end{cases}$	$\begin{cases} 6_3 \\ 6_3^3=3_3 \end{cases}$	$\begin{cases} 1 \\ 3 \end{cases}$	$\begin{matrix} H3_2 \\ H6_2 \end{matrix}$
	$P(xy0)$	2	$6_3^3=2$	4	
$H6_4$	$P(xy\frac{1}{6})$	$\begin{cases} 6 \\ 3 \end{cases}$	$\begin{cases} 6_4 \\ 6_4^3=3_1 \end{cases}$	$\begin{cases} 1 \\ 3 \end{cases}$	$\begin{matrix} H3_1 \\ H6_4 \end{matrix}$
	$P(xy0)$	2	$6_4^3=2$	4	
$H6_5$	$P(xy\frac{1}{6})$	$\begin{cases} 6 \\ 2 \end{cases}$	$\begin{cases} 6_5 \\ 2_1 \end{cases}$	$\begin{cases} 1 \\ 4 \end{cases}$	$\begin{matrix} P2_1 \\ H6_5 \end{matrix}$
	$P(xy0)$	3	$6_5^3=3$	3	

Crystal

Harker

Implication

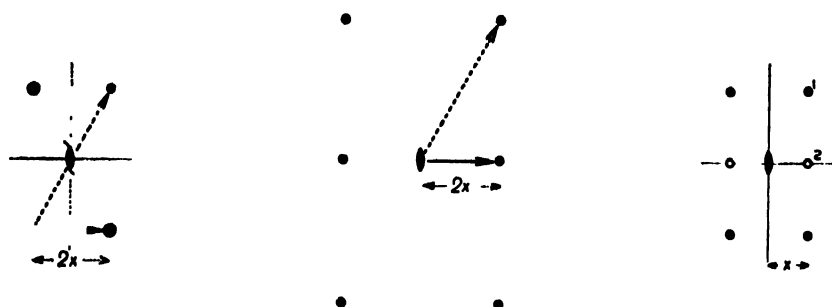


FIG. 7. The appearance of satellitic peaks due to reflection symmetry. The relative heights of the peaks are indicated by small numerals on the implication diagram.

ghost: $R3$, $R32$, and all the axial isometric space groups, namely $P23$, $F23$, $I23$, $P2_13$, $I2_13$, $P43$, $P4_33$, $F43$, $F4_33$, $I43$, $P4_33$, $P4_13$, and $I4_13$. Further useful information of this character is assembled in Table IV.

The conclusions of this section can be summarized as follows: For 18 of the 21 possible parallel-axial symmetries (i.e., all symmetries except $P4$, $I4$, and $H6$), there exists an implication diagram which is a map of the crystal structure repeated m ($=4, 3, 2$, or 1) times by translation (and, in the case of trigonal crystals lacking a center of symmetry, repeated again by inversion). For each axial space group containing one of these parallel-axial space groups as subgroups, therefore, there exists an implication diagram which is a map of the crystal structure plus $m-1$ ghosts (all repeated by inversion in the case of trigonal crystals lacking a center of symmetry). For the three axial symmetries $P4$, $I4$, and $H6$, all possible implication diagrams contain, in addition, one or more times this number of satellitic peaks due to subgroup operations.

Satellitic Peaks Due to Reflection Symmetries

The addition of reflection symmetry in planes normal to the symmetry axes does not give rise to any additional peaks. The addition of symmetry planes parallel to the symmetry axes may or may not give rise to satellitic peaks on implication diagrams. For purposes of discussing this, suppose that all symmetry axes are regarded as screws (the screw component of which may be

zero in the special case of a pure rotation axis) and that all symmetry planes are regarded as glide planes (the translation of which may be zero in the special case of a pure reflection plane). Suppose that the Harker synthesis,

$$P\left(xy - \frac{1}{w}\right),$$

is made for a translation component, $1/w$, of the screw. Then, if the symmetry plane does not also have a translation component, $1/w$, it does not provide additional atoms separated by this level difference. Hence it does not give rise to new interatomic vectors of this separation, and

TABLE IV.

Symmetries without satellitic peaks			
Total ambiguity factor	Centro-symmetry factor	Translation ambiguity factor, m	Parallel-axial groups
1	1	1	$R\bar{3}$, $H6_1$, $H6_2$
2	2	1	$R3$
2	1	2	$C2$, $I\bar{4}$, $P4_1$, $P4_2$, $P4_3$, $I4_1$
3	1	3	$H3$, $I\bar{3}_1$, $H3_2$, $H6_3$ †
4	1	4	$P2$, $P2_1$, $P\bar{4}_1$ †, $H6_4$ †, $H6_5$ †
Symmetries with satellitic peaks			
Number of sets of satellitic peaks	Translation ambiguity factor, m	Parallel-axial groups	
1	1	$I4$, $H6_6$ *	
1	2	$P4$	
2	1	$H6_6$ *, $H6_7$ *	
3	1	$H6$	

* Subgroup is without satellitic peaks.

† Subgroup only.

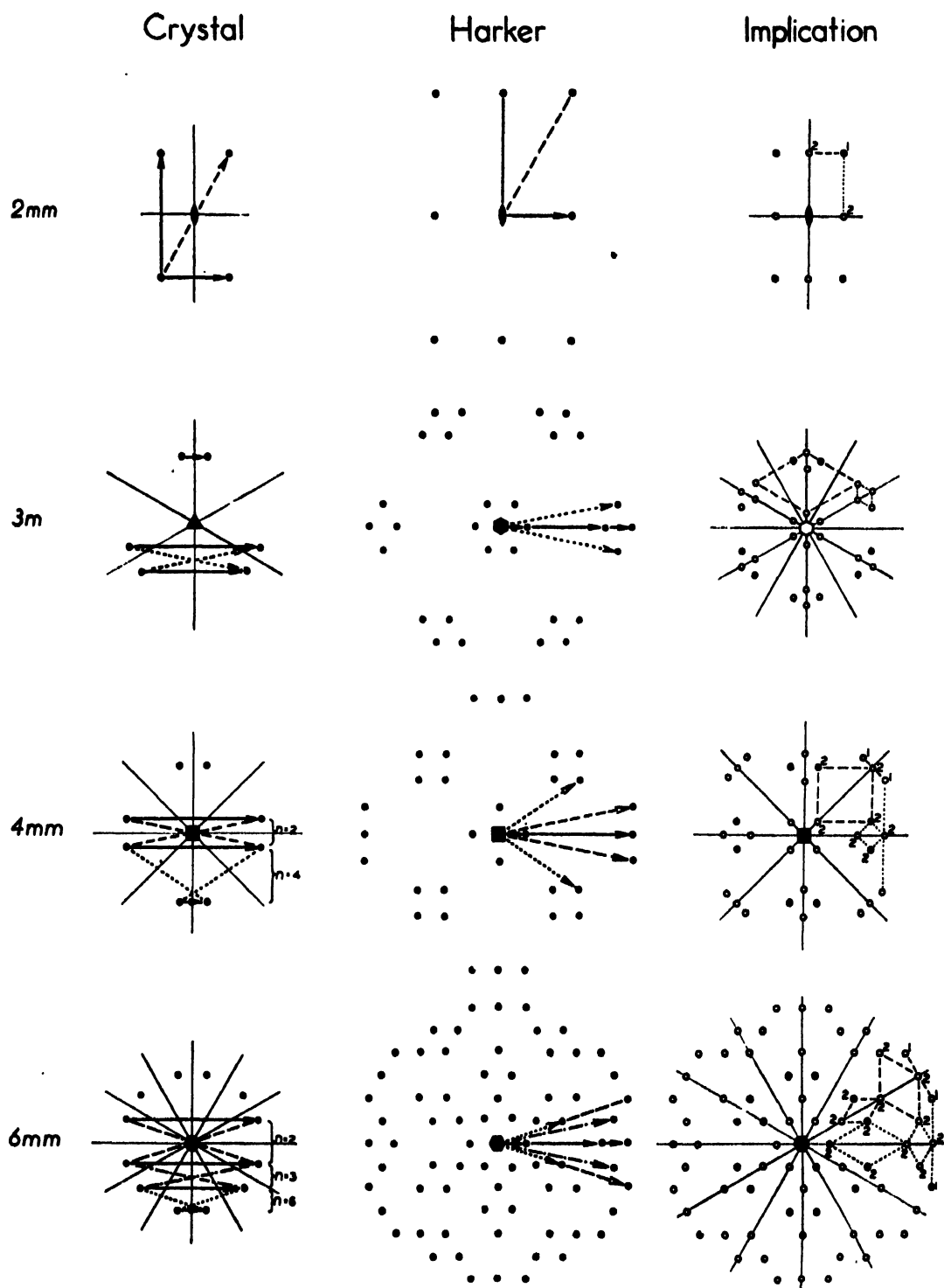


FIG. 8. The relations between peaks and satellitic peaks for the symmetries $2mm$, $3m$, $4mm$, and $6mm$. The relative heights of the several peaks, where these are different, are indicated by small numerals on the implication diagram. Points and symmetry not present in the crystal are indicated by outlined symbols.

consequently no new peaks appear on the Harker synthesis,

$$P\left(xy - \frac{1}{w}\right).$$

In this case the implication diagram is a map of the crystal structure, with the exception of the ambiguities and satellitic peaks already discussed. Important cases of this are two-, three-, four-, and sixfold axes combined with parallel glide planes whose glide component is $c/2$.

If, on the other hand, the translation component of the screw axis and the axial component of the glide plane are the same, the reflection symmetry introduces additional atoms at the same level, and the Harker syntheses contains peaks representing vectors from each atom both to its rotation and its reflection equivalent. It is convenient to divide the discussion of such instances into two cases, according as the plane contains the axis or is parallel with the axis.

Symmetry Plane Containing Symmetry Axis

Figure 7 shows the simplest example of this case, namely a rotation and reflection combination in which only one plane has the same translation component as the axis. This example may serve as a basis for the development of more complicated examples. The solid circles and double circles each constitute sets which, in themselves, are related by the twofold screw only. The interatomic vectors for these sets within themselves are the usual ones, and are shown in the general positions of the corresponding Harker and implication diagrams. There is only one other kind of interatomic vector introduced by glide plane, namely a vector at right angles to the plane; this is shown in the lower part of the crystal diagram. Such a vector always gives a point in a specialized location on the Harker and implication diagrams, namely on the central line at right angles to the plane. Its coordinates are $2x, 0$ on the Harker diagram for the situation shown, but $x, 0$ on the implication diagram. Since the relation of scale on implication to Harker varies with n , it is better to observe the general geometry of the location of this satellitic point with respect to the true implication points than merely the coordinates of the satellitic points.

In Fig. 8 are shown the relations between the atoms in the crystal, the Harker diagram, and the implication diagram for the important cases of reflection planes containing the several n -fold rotation axes. The Harker diagram in each case is derived in a manner similar to the simple case just treated: The atoms are first regarded as divided into two axially symmetrical sets. The two sets are related to one another by the several reflection symmetries of the planes. The total number of interatomic vectors in the two sets consists of (a) the vectors within each set due to axial symmetry, plus, (b) the vectors from one set to the other due to each reflection plane. The first collection of vectors, (a), locates, in the implication diagram, just the positions of the atoms in the structure (plus attendant ambiguities characteristics of n , and satellitic peaks previously discussed). The second set of vectors all lie in special positions, namely on lines in the Harker diagram at right angles to the plane producing the vector in the crystal. In the implication diagram, this line is rotated out of normality with the plane producing the points by an amount characteristic of the number n (see Table II). This rotation places the line normal to another plane of the centrosymmetrical group. The peaks at the ends of these vectors on the implication diagram do not represent possible positions of atoms. They can be recognized by their geometrical relation to the peaks in the general position. In the diagrams of Fig. 8, the derivation of the satellitic peaks on the Harker diagram due to a single reflection plane of the symmetrical set is given. Vectors represented by solid lines are those due to reflection. Vectors represented by broken lines are certain ones due to axial symmetry which are included to show the relation of rotation vectors to reflection vectors in producing the geometry of the pattern of true and satellitic peaks. Note that for $n > 3$, both axial and reflection vectors occur for powers of operations. These are shown in bracketed sets in the left column of Fig. 8.

The method of analysis of diagrams having axes with intersecting reflections is indicated in the right column of Fig. 8. To relate satellitic peaks to non-satellitic peaks caused by the same symmetrical set of atoms, first locate the non-

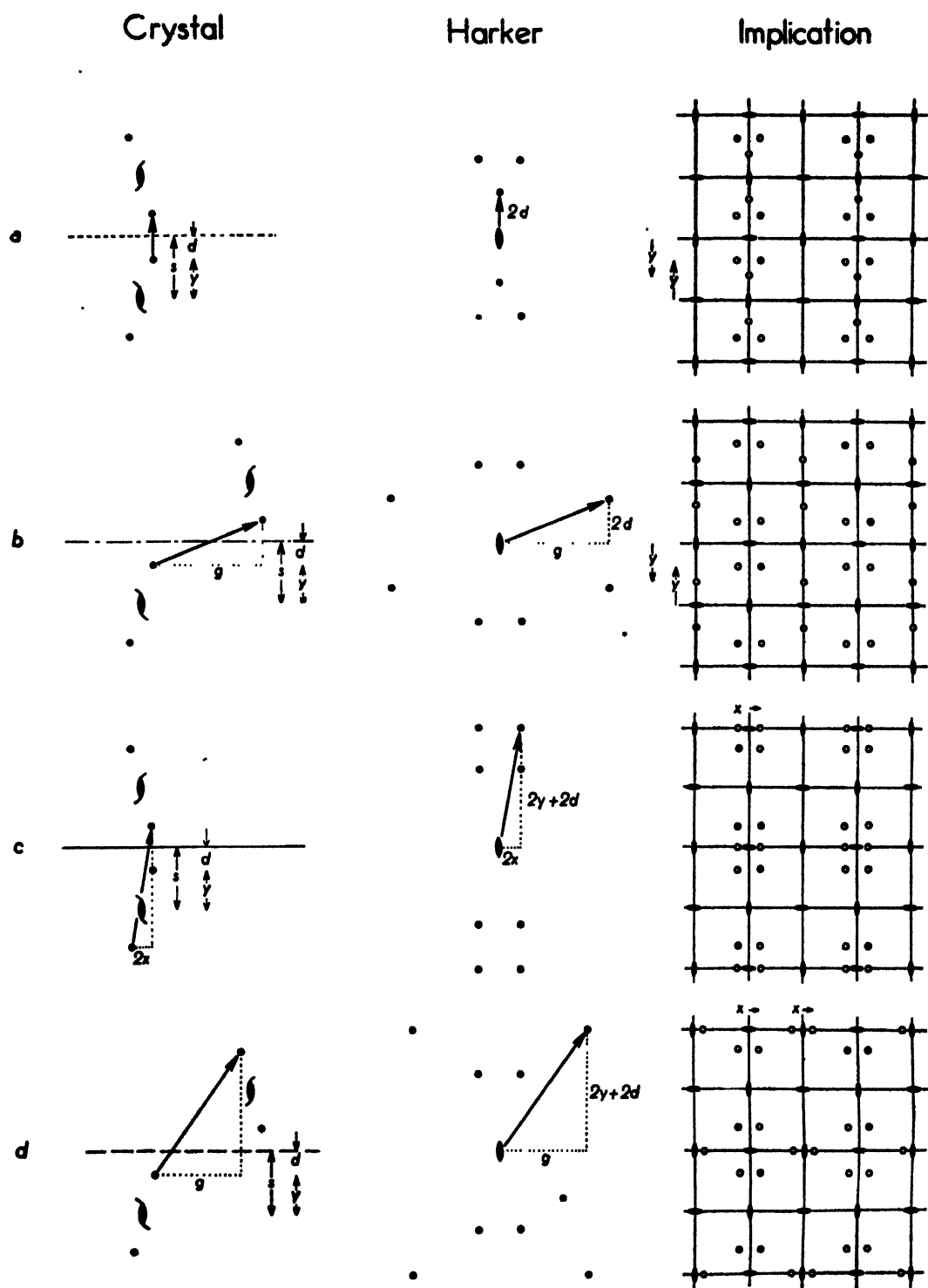


FIG. 9. The possible combinations of twofold screws with parallel reflections, and their implication diagrams. Points and axial symmetry in the implication diagram which have no counterparts in the crystal are shown by outlined symbols.

satellitic peak. Unless the atom is in a special position on one of the reflection planes, this is the peak in the general position closest to the origin. From this peak, drop a pair of lines to each of the two nearest symmetry planes of the crystal (for $n=3$, ignore the spurious extra planes introduced by centrosymmetry; they can be identified by not containing satellitic peaks) and making angles with the plane of 90° for $n=2$, $\pm 60^\circ$ for $n=3$, $\pm 45^\circ$ for $n=4$, or $\pm 30^\circ$ for $n=6$. (For $n=2$, these two lines coincide, causing a satellitic peak of double height.) At the points of intersection of the two lines with each symmetry plane lie the two satellitic peaks due to reflection for $n=2, 3, 4$, or 6 , respectively.

This completes the process for $n=2$ or 3 , but for $n=4$ or 6 , there are further satellitic peaks due to powers of the operation and their attendant reflections. The following directions for analyzing these additional peaks rest on the fact that (left column of Fig. 8) for the fourfold case, a reflection vector is common to the two zones $n=4$ and $n=2$, while for the sixfold case, a reflection vector is common to the two zones $n=6$, $n=3$, and another is common to the two zones $n=3$, $n=2$. Thus the outer reflection peaks just found are also related to the required axially satellitic peaks. For the fourfold case, draw a line at right angles to each of the two non-equivalent reflection planes at the positions of the outer satellitic peaks, found as described above. These two lines intersect each other in the axially satellitic peak. This completes the analysis for the fourfold case. For the sixfold case, the first subgroup has $n=3$. To find the axially satellitic peak corresponding with this, drop lines from the two outer reflection-satellitic peaks (one on each plane) as found above, each line making an angle of 60° with its plane, the lines converging outward. These intersect in an axially satellitic peak due to the subgroup with $n=3$. Complete the $n=3$ diamonds, thus locating the $n=2$ reflection-satellitic peaks at their outer points. At these draw perpendiculars from the planes. Their intersections are the $n=2$ axially satellitic peaks. This completes the process for the sixfold case.

In this manner, starting with a peak in a general position, all satellitic peaks caused by the same symmetrical set of atoms can be located.

Thus located, satellitic peaks can be discarded and only non-satellitic peaks need be considered in the further analysis of the implication diagram.

By giving attention to the geometrical relations between rotation and reflection peaks on implication diagrams, such diagrams can thus be easily interpreted for cases where the axis contains the plane. The only theoretical difficulty arises when the atoms responsible for the peaks are not in the general position but rather are located on the specialized position on the line through the axis and normal to the plane. When this occurs, the peak due to such an atom is correctly located on the implication diagram, but the location superficially suggests that the peak is satellitic. It can be distinguished from the reflection-satellitic peaks just discussed because it is not geometrically related (in the manner discussed) either to (a) peaks in the general position, or (b) to other peaks on its own symmetry line.

Symmetry Plane Parallel with Axis

Even when the symmetry plane does not contain the axis, the positions of reflection-satellitic peaks are confined to specialized locations so that these can be recognized. Several simple cases involving twofold screws and representative kinds of reflections are shown in Fig. 9. Two different kinds of cases can be recognized. In general, the atoms of a set $a_1, a_2, a_3 \dots$, are related by the operations of rotation, and this set is related to another set, $a_{1s}, a_{2s}, a_{3s} \dots$, by the operation of reflection. If the Harker synthesis is carried out (symmetry permitting) so that vectors $1-1s, 2-2s, 3-3s \dots$ result, then the reflection-satellitic peaks are confined to a line normal to the reflection plane (Fig. 9a and 9b), a situation similar to that resulting when the reflection plane contains the axis, as discussed in the last section.

A slightly different situation occurs when the Harker synthesis is carried out so that vectors other than $1-1s, 2-2s \dots$ result. Figs. 9c and 9d illustrate cases containing twofold axes where vectors $1-2s, 2-1s$ result. This general situation can be analyzed as follows: Let the operation of rotation relate a_1 to a_j , of which the corresponding atoms by reflection in the plane are a_{1s} and a_{js} . Let the Harker synthesis be car-

ried out to that vector $j-1s$ results. Now a_j is related to a_{1s} by a combination of a rotation (to a_1) followed by a reflection (to a_{1s}). This combination is equivalent to a reflection in some other plane of the space group making an angle with the first of $\frac{1}{2}(a_1 \wedge a_j)$. Thus the vector $j-1s$ is normal to this plane, and the reflection-satellitic peak again appears in the implication diagram on a line normal to some reflection plane in the space group.

In both the above instances, if the glide of the reflection operation has a component normal to the axis of amount g (Fig. 9b and 9d), then the reflection-satellitic peaks are still on a line normal to the reflection plane, but are displaced by an amount $g/2$ from the axis.

INTERPRETATION OF IMPLICATION DIAGRAMS

The conclusions arrived at in the foregoing discussions may be summarized in more compact form as follows: Implication diagrams may be prepared for any space group involving axial symmetries with or without additional reflection symmetries. The diagram is a map of the crystal structure, usually, but not always, complicated by the presence of peaks not corresponding with any atoms in the crystal structure. These unwanted peaks may be divided into two categories, namely ambiguities and satellites.

The ambiguities are, in effect, ghost images of the crystal structure repeated by inversions (in the case of threefold axes only), and translations (not the translations of the space group). The second kind of ambiguity is a consequence of the translational symmetry of crystals. The number of ghosts and their placement, arising from this cause, is a function of the number of operations in the axial symmetry group. It is usually impossible to resolve this ambiguity.

Satellitic peaks differ from ambiguities in that they are not ghosts. They invariably occur in positions where they can be recognized and rejected. They are of two kinds: rotation-satellitic peaks and reflection-satellitic peaks. The rotation-satellitic peaks occur in a special pattern, geometrically related to non-satellitic peaks. The reflection-satellitic peaks are concentrated in lines normal to reflection planes, and also occur in recognizable patterns.

Satellitic peaks are missing in space groups

characterized by purely axial symmetries of certain sorts. They are absent when the Harker synthesis can be made for the following axial symmetries: 2, 2₁, 3, 3₁, 3₂, $\bar{3}=3+2$, 4, 4₁, 4₂, $\bar{4}$, 6, 6₁, and $\bar{6}=3/m$. They are also absent for certain Harker syntheses of 6₂, 6₃, and 6₄. They are also absent for certain axial symmetries combined with glides having translation components parallel with the axes. For all such cases, the interpretation of implication diagrams is comparatively simple, for the implication diagram is the crystal structure plus $m-1$ ghosts. (The value of m is given in Table I.)

The value of m is unity for sixfold axes. As a consequence of this, space groups containing the two kinds of sixfold screws listed in the last paragraph are without ghosts. For this reason, the implication diagram is a map of the crystal structure for the particular space groups $H6_1$, $H6_2$, $H6_3$, and $H6_4$ (C_6^2 , C_6^3 , D_6^2 , and D_6^3 , respectively). This is also true of $R\bar{3}$, except that the implication of this group has a Harker peak at the origin. This opens the possibility of solving the structures of crystals belonging to such space groups, no matter how large the cell.

In addition to the ambiguities and satellitic peaks, each Harker synthesis for $P(xy0)$ has a satellitic peak at the origin. This is not true for syntheses

$$P\left(xy\frac{1}{w}\right),$$

where $1/w$ is different from zero. For this reason, other things being equal, the latter synthesis is to be desired. If it is desirable for other reasons to utilize $P(xy0)$, then the additional synthesis

$$P\left(xy\frac{1}{w}\right)$$

should also be carried out if possible merely to locate a possible atom at the origin.

SPACE GROUP DETERMINATION BY HARKER SYNTHESIS

It is generally recognized that ordinary x-ray diffraction studies do not provide criteria for distinguishing the 230 space groups from one another in all cases. Rather, one is able to distinguish 121 diffraction groups⁵ from one an-

⁵ M. J. Buerger, *X-Ray Crystallography* (John Wiley and Sons, New York, 1942), p. 510-511.

other. Where a diffraction group contains only one space group, then that space group is uniquely characterized, otherwise not. This is fundamentally because such methods depend on extinctions, which are able to detect symmetries with translation components only. Thus it is impossible to distinguish between 1, 2, and m , between 4 and $\bar{4}$, or between $H3m1$ and $H31m$. For example, within the diffraction group $mmmP$ — — —, the same extinction effects and other qualitative diffraction effects are given by $P222$, $Pmm2$, and $P2/m\ 2/m\ 2/m$.

Harker synthesis may be described as quantitative x-ray crystallography, since by its aid it is possible to distinguish all space groups except pairs differing only by an inversion. In other words, by means of Harker syntheses, it is possible to push x-ray crystallography to the limits set by the fact that diffraction effects are centrosymmetrical. This is because a Harker synthesis provides unique characterizations of the symmetry elements m , 2, and 4 *vs.* $\bar{4}$, as well as of other symmetry elements. This is an outcome of the fact that x-ray data is utilized in a quantitative way as against the qualitative criterion based on the mere presence or absence of reflections used in ordinary x-ray crystallography. Harker synthesis provides these space group criteria without any attempt being made to interpret the synthesis for the locations of the atoms.

To give an example of the way the synthesis can be used in space group identification, suppose one has a crystal belonging to the diffraction group $mmmP$ — — —. The possible space groups are $P222$, $Pmm2$, and $P2/m\ 2/m\ 2/m$. The distinction between these can be settled on the basis of reflection-satellitic peaks. Representing the presence of satellitic peaks by S and their absence by 0, the Harker syntheses for the several axes give the following combinations of satellitic peaks:

$P222$	000
$Pmm2$	SS2
$P2/m\ 2/m\ 2/m$	SSS

It is even easier to distinguish between such cases as $H3m1$ and $H31m$. The distinction between these two depends on whether the symmetry planes are parallel to the long diagonal or the short diagonal of the cell. In Fig. 8, the right

column shows that the threefold case has symmetry planes parallel to both directions, *but the true symmetry plane is marked on the implication diagram by containing the satellitic peaks.* The false symmetry planes lack the satellitic peaks. (On the Harker diagram the situation is reversed.) Since the long and short diagonals are easily distinguished on the implication diagram, this analysis quickly fixes the space group.

NON-HARKER BACKGROUND

Although Harker pointed out³ that certain one-dimensional and two-dimensional Patterson syntheses could be used to find interatomic vectors between atoms not related by symmetry, it will nevertheless be convenient to adopt the following nomenclature:

Harker peak: a peak corresponding to an interatomic vector between symmetry-equivalent atoms.

Non-Harker peak: a maximum corresponding to a vector between atoms or parts of atoms which are not symmetry-equivalent.

One of the confusing aspects of Harker and implication diagrams is that they contain a non-Harker background. This is always present in varying degree. Its characteristics may be such as not to seriously affect the interpretation of the diagram, or, on the other hand, it may furnish peaks of magnitude equal to, or greater than, the Harker peaks. Since one does not usually know at the outset of the analysis of a crystal structure which of these cases he is dealing with, it is dangerous to assume that every peak occurring on a Harker or implication diagram is a Harker peak. This point is emphasized here because there appears to be a prevailing opinion that any peak on a Harker diagram is a Harker peak; or, if non-Harker peaks are recognized at all, that at least all strong peaks are Harker peaks.

Non-Harker background would be absent, except for special cases such as mentioned beyond, if atoms were composed of matter concentrated at discrete points. The background arises from the circumstance that the electronic matter of atoms is distributed over a comparatively large volume. For pairs of symmetrically equivalent atoms, every point in one atom is separated from a corresponding point in the other atom by the translation component of the symmetry operation. But there are usually parts of the volumes

of pairs of atoms not related by symmetry which are separated by the same translation. The integration of the Patterson function extends over the entire height of symmetry-equivalent atoms, but only over as much of the height of non-equivalent atoms as overlap by the translation component of the symmetry element. This effect, of course, usually causes the peak due to the interaction between non-equivalent atoms to be comparatively weak compared with that due to equivalent atoms. The difference is augmented by the fact that the dense centers of equivalent atoms interact, while in the case of a pair of non-equivalent atoms either the diffuse peripheries interact, or at best, the dense center of one interacts with the diffuse periphery of the other, unless the atoms are fortuitously separated by exactly the translation component of the symmetry element.

Yet non-equivalent atoms can give rise to strong peaks. For example, suppose that a very heavy atom is on the symmetry axis. Its Harker representation may be a peak at the origin which is not distinct from the origin peak. The Harker diagram selected for this axis will contain strong non-Harker peaks if any other atom has a coordinate differing from that of the heavy atom by the translation component, $1/w$, of the synthesis, and the peak will also be strong if its coordinate has approximately this value.

Furthermore, in simple structures, or in structures derived by substituting different atoms in simple structures (substitution structures), pairs of atoms not symmetrically related may occupy levels differing by 0 , $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, or $\frac{1}{6}$, which are the level differences utilized for finding interatomic vectors between symmetry-equivalent atoms for various axes. In these instances, intense non-Harker peaks appear on the Harker diagrams. While it is not generally possible to distinguish between Harker and non-Harker peaks, the latter should be suspected in crystals of simple structure and in crystals expected to show substitution relation to simpler structures, for example superstructure crystals. In syntheses involving satellitic peaks, the non-Harker peaks are not accompanied by satellitic peaks located in the same way as those to be expected for the particular implication. In these cases, non-Harker peaks can be recognized.

HEIGHTS OF HARKER PEAKS

It will be recalled that the value of the three-dimensional Patterson function at a point, $P(xyz)$, is the average value of the product of the electron densities at the ends of a vector defined by a line extending from the origin to the point xyz , as the vector, fixed in direction and length, is allowed to range over the volume of the cell. For a pair of atoms, A , related by symmetry, the two atoms at opposite ends of the vector, are the same, and the height, H , of the Harker peak, above any non-Harker background, is given by

$$H_A = (\rho_A^2)_{\text{cell } w} \quad (6)$$

where $(\rho_A^2)_{\text{cell } w}$ is the average squared density of the atom A , the average being taken over the cell. Now, the average over one volume differs from the average value over another volume merely by a scale constant. It is possible to determine $(\rho_A^2)_v$, the average squared density of the atom over an arbitrary volume, v , provided only that the arbitrary volume is greater than that of any atom which might come up for consideration. For any cell, the peak height, H_A , due to the Harker interaction between a pair of identical atoms can then be expressed

$$H_A = k(\rho_A^2)_v \quad (7)$$

The proportionality constant holds for all atoms in the cell, but changes when a new cell is considered. The value $(\rho_A^2)_v$ is characteristic of an atom, and the interpretation of Harker diagrams would be aided by a list of such values for the several chemical atoms.

The function (7) has a high value for compact atoms and a small value for diffuse atoms. Consequently, Harker peaks tend to be high for the compact cations and low for the diffuse anions. Thus, for the silicates, the silicon and oxygen atoms, which are approximately isoelectronic, provide Harker peaks between which there is great contrast. Indeed the oxygen peaks are swamped by background. Peak contrasts in such cases are even more pronounced than the peak heights would lead one to believe, for the background contrast is enhanced by the fact that atoms with large volumes have peaks which decline gradually over larger areas than peaks due to small atoms.

Dissociation Energies of Surface Films of Various Oxides as Determined by Emission Measurements of Oxide Coated Cathodes*

HAROLD JACOBS

*New York University, New York, New York***

(Received March 11, 1946)

A mechanism of falling emission in vacuum tubes is discussed. It is demonstrated that when an electron achieves a critical kinetic energy in moving from the cathode to the anode, and if the anode is an oxide, the electrons will cause a dissociation of the oxide. The liberated oxygen will return to the cathode and falling emission will result. The critical energy of the electron starting the dissociation is found to be equivalent to the heats of formation of the oxides bombarded in the case of five different oxides. Using this equivalence principle, the heats of formation of two compounds are found which have not been recorded in the literature, Ta_2O_4 and ZrO . In addition, the contact potential between barium oxide cathodes and seven various oxides are determined and the work function of the seven oxides computed.

PART I

Introduction

FOR some years it has been well known that, when a metallic oxide, such as nickel oxide or copper oxide, is bombarded by electrons, this oxide will decompose releasing large quantities of gas. If such bombardment of an oxide takes place in a vacuum tube, this may result in "gassy" tubes and "low emission" tubes.¹ Usually, the emission in the tubes will noticeably decrease and the gas pressure may increase. This falling emission will be shown in a series of experiments later in this paper. Although it has been thought that oxides decompose upon electron bombardment, there has not been very much reported as to the fundamental manner by which the phenomenon of the release of gases from oxides upon electron bombardment occurs.

In this report, we shall propose an hypothesis, and then subject this hypothesis to experiment. An electron in motion will have a certain kinetic energy. When this electron bombards a surface film of oxide it transfers this energy to the bombarded material. As the kinetic energy of the electron reaches a critical value, the oxide will decompose. We propose that the critical energy of the electron required for decomposition of the oxide corresponds to the heat of formation

of one molecule of the oxide. We shall determine the energy of decomposition by bombarding the oxide at various applied voltages and observing emission from an oxide coated cathode. As the bombarded material emits gases, these gases "poison" the oxide coated cathode and falling emission is observed as a function of bombarding voltages.

If a relationship can be shown between the heats of formation of molecules and the energy required for dissociation, we shall be able to determine the heats of formation of molecules whose values have not been determined by physical chemical methods. However, before discussing the hypothesis further, we shall discuss in some detail the nature of alkaline earth oxide cathodes, since this is the tool used throughout all of these experiments.

A great deal of work has been done on the alkaline earth oxide cathodes during the nineteen-twenties and nineteen-thirties. It was found that because of an extremely low work function these materials produced a copious electron emission at relatively low temperatures.² These cathodes are usually composed of physical or chemical mixtures of barium oxide and strontium oxide ($BaO + SrO$). Sometimes, another mixture is used containing a third component in addition to the other two, calcium oxide (CaO). In the rest of this paper, we shall use barium in the written work, but it should be remembered that the properties of calcium and strontium can be

* Summary of thesis submitted in partial fulfillment of the requirements for the degree of doctor of philosophy at New York University, July 10, 1945.

** The author is now employed at Sylvania Electric Products, Inc., Kew Gardens, New York.

¹ L. B. Headrick and E. A. Lederer, *Phys. Rev.* 50, 1094 (1936).

² A. Wehnelt, *Ann. d. Physik* 14 (1904).

described throughout in a somewhat similar fashion. We choose only to speak of barium since this is the component which has the lowest work function³ and is most often referred to in the literature.

To make an indirectly heated cathode, the alkaline earth carbonates are often sprayed on a nickel sleeve and mounted in a tube. During exhaust the cathodes go through a process known as "activation." The cathodes are heated to about 1000°C and the carbonate decomposes according to the reaction



The carbon dioxide is pulled out through the exhaust tubing leaving a deposit of barium oxide on the nickel sleeve. Upon further heating and application of voltage to tube elements, free barium is formed, some of which diffuses to the surface of the cathode in a process which will be discussed later. It is the presence of this free barium which is responsible for the low work function and high electron emission from the cathode.⁴ In the "activation" of a cathode, there are many processes at work, all of which combine to form a sort of equilibrium.

Some of the factors at work are as follows. On the side of those factors producing free barium, we have

- (a) reduction by core metal,⁵
- (b) electrolysis of the cathode coating,⁶ and
- (c) reduction of the cathode by gases.⁷

Several factors have been indicated by experiments as hindering the production of free barium.

- (a) High voltage operation has brought about a loss of surface barium in the form of sputtering by ions formed from residual gases in the tubes.⁸

³ J. P. Blewett, Gen. Elec. Bull. No. 1001 (December, 1931).

⁴ J. A. Becker and R. W. Sears, Phys. Rev. **38**, 2193-2213 (1931).

⁵ E. F. Lowry, Phys. Rev. **35**, 1367-1378 (1930).

⁶ J. A. Becker, Phys. Rev. **34**, 1323-1351 (1929).

⁷ J. H. DeBoer, *Electron Emission and Absorption Phenomenon* (Cambridge University Press, Cambridge, England, 1935), p. 365.

⁸ A. W. Hull, Gen. Elec. Rev. **32**, 213 and 390 (1929). Massachusetts Institute of Technology Department of Electrical Engineering, *Applied Electronics* (J. Wiley & Sons, New York, 1943).

- (b) Residual gases have been found chemically to combine with free barium. Furthermore, the barium oxide is quite permeable to these gases⁹; i.e.,



- (c) Some free barium is continually being lost by evaporation.¹⁰

Thus we have a balance of factors, the slightest disturbance of which can cause large variations in emission. This balance was the indicator used in detecting at what voltages dissociation of oxides started to occur.

PART II

Method of Determining Critical Kinetic Energy of Electrons Required for Falling Emission to Occur

A. Vacuum Tubes

Tubes were made as per sketches in Fig. 1. These tubes consisted of a Nonex glass envelope with tungsten leads in which a rotating nickel cylinder that could be turned by a magnet outside the tube was assembled in the mount structure. On this cylinder was welded a strip of the metal to be investigated. This strip was as long as the cylinder and somewhat wider than the cathode. Great pains were taken to insure purity and lack of contamination of the strip.

The metal strips chosen for test purposes were taken from metals such as are used for spectroscopic standards. All free metal parts were acetone washed, rinsed in distilled water, acid washed, distilled water washed, acetone-rinsed again, and hydrogen fired for ten minutes at 800°C. During sealing, N₂ flushing gas and low fires were used to prevent oxidation.

B. Exhaust

The tubes were then sealed on an exhaust system and the metal was oxidized at exhaust.

⁹ J. A. Becker, Phys. Rev. **34**, 1323 (1939). Several tubes were made by the author containing an oxide coated cathode and a square anode (1 cm × 1 cm) a distance of $\frac{1}{4}$ cm from the cathode surface. These tubes were placed on a vacuum system and emission measurements made. In the first tube with the pressure < one micron, the emission was 12 μ a at 5 v on the anode and 650 μ a at 30 v. In the presence of 15 μ of oxygen, the emission dropped to 0 μ a at both 5 v and 30 v. In the second tube, the emission read 15 μ a and 360 μ a at a pressure < one micron. In the presence of three microns of oxygen, the emission read .3 μ a at 5 v and 1.9 μ a at 30 v. All measurements were made at 7.0 v, .66 amp. on a 6.3-v filament. At lower wattages on the cathode filament the effect of oxygen was found to increase.

¹⁰ L. R. Koller, Phys. Rev. **25**, 671-676 (1925).

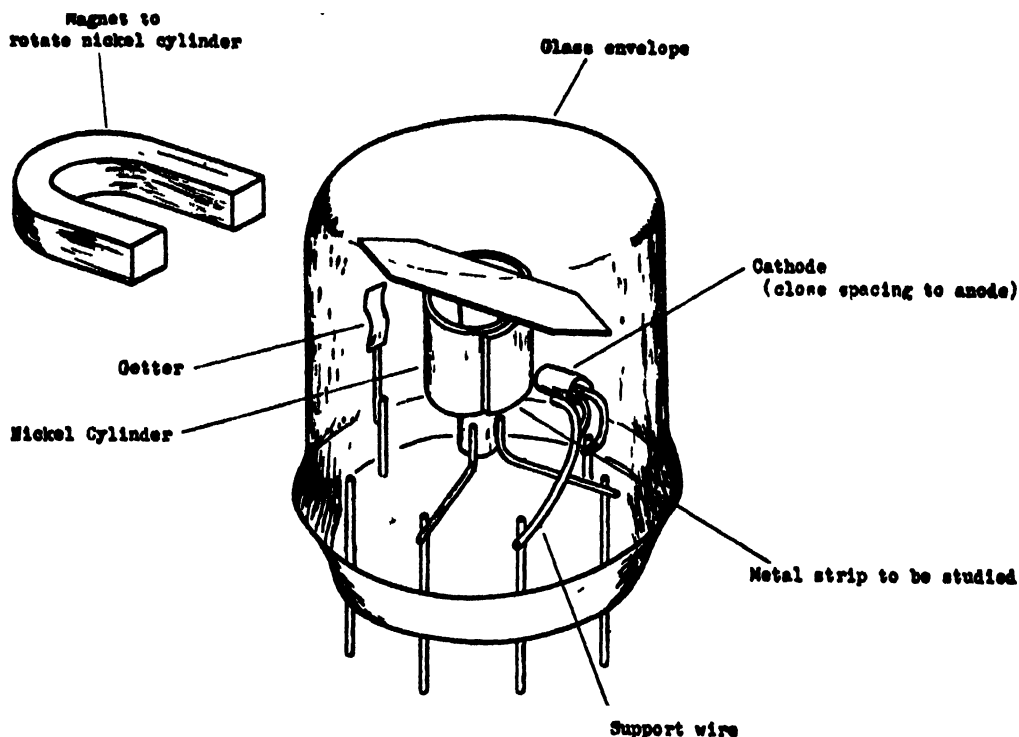


FIG. 1. Experimental tube for measurement of dissociation energies of surface films of oxides

This exhaust system contained a mercury cut-off and an AgO flask. A high frequency furnace was used to heat treat the anode for 30 seconds at 800°C brightness in 500 μ of oxygen.

C. Method of Measurement

The filament in the cathode sleeve was usually adjusted to roughly about half the rated voltage, i.e., 3.5 v, .38 amp instead of 6.3 v, .65 amp. This would bring the temperature down to an estimated temperature region of 350° to 450°C.

The metal oxide to be studied had been kept free of contaminations from the cathode activation and aging by rotating it away from the cathode. After the cathode temperature was lowered, the sample was rotated by means of a magnet to a position directly in front of the cathode. The use of low cathode temperatures was essential because

- (a) metallic evaporation from the alkaline earth cathode could be kept to a minimum, and
- (b) the effect of oxygen in inhibiting emission was more marked with lower temperatures of the cathode.

A voltage was then applied to the anode and increased in $\frac{1}{2}$ -volt steps for one minute each.

At lower voltages, the current would rise with the voltage and not decay with time; however, at a critical voltage, the emission would slowly start to fall. Beyond this voltage, the decay in emission became more marked. This critical voltage was noted.

By varying filament currents, the cathode emission could be varied. It was found that falling emission is independent of the wattage but is only dependent on electron volts of energy applied. Furthermore, time was not important, for below the critical voltage, the emission could be maintained.

Now the voltage applied is not the true measure of energy of the electron because of contact potential and thermal kinetic energy of the electrons. The true energy of electrons was approximated as follows. The measurement of falling emission was made. This was determined by the beginning of falling emission as recorded by a milliammeter. The smallest amount of falling emission which could be determined by this instrument was in the order of one microampere. Therefore, the anode was biased down to that voltage (E_p) on the anode such that emission

from the cathode of one microampere was just reached. This latter voltage was arbitrarily called the retarding cut-off potential (E_o). Then $E_p - E_o$ was used as the real potential (Fig. 2).

$$E_p - E_o = E \text{ real.} \quad (3)$$

These measurements were repeated several times at different wattages and on several samples of similar types. In addition to this, a further check was made on sample to insure that the same surface was being bombarded. This was done by measuring the contact potential between each surface and the cathode. If the contact potentials were the same from tube to tube of the same sample, the consistency of producing the same oxide surface could be further assured.

The method of measuring the contact potential was done as follows. The filament voltage was still further lowered and both accelerating and retarding potentials were applied, emissions measured and extrapolated. Against this, accelerating potentials were applied and extrapolated (Fig. 3).

The point at which these extrapolations meet was assumed to be the contact potential, for if there were no contact potential, the retarding potential would change to an accelerating potential at 0 volt (Fig. 4).

Thus the shift in the theoretical curve is caused by a contact potential E_{contact} between the work function of the barium cathode ϕ and the work function of the oxide being studied ϕ' , or in other terms, the difference in the work functions of the two materials

$$\phi - \phi' = E_{\text{contact}}. \quad (4)$$

PART III

Data

The data in Table I were taken on tubes,

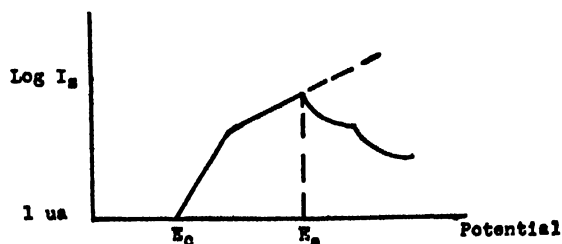


FIG. 2. Measurement of electron energy. I_e —emission, E_a —applied potential, E_o —cut-off potential.

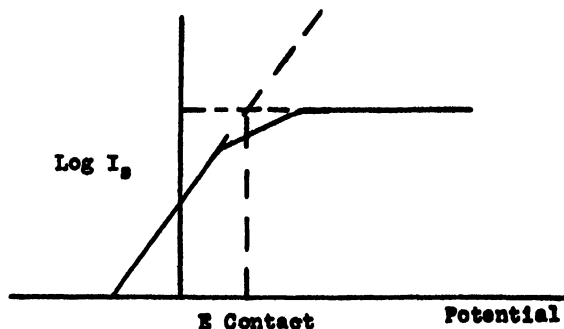


FIG. 3. Shift in emission due to contact potential.

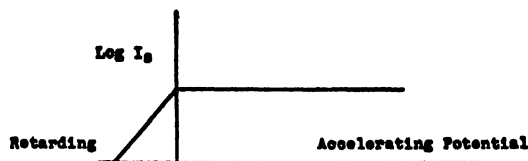


FIG. 4. Ideal emission curve with no contact potential.

anodes of which were prepared as in the discussion above. It might be pointed out that the values "critical potential" for falling emission to occur were independent of time, distance, and current density of the bombarding current provided that the anodes were not caused to increase in temperature to any great extent.

Figure 5 illustrates a typical case of how falling emission occurs as the voltage is raised. In this case the voltage was raised in 0.5-volt steps and held at each level for three minutes. The particular sample being studied under bombardment was FeO.

In the study of falling emission while bombarding oxides with electrons at various voltages, the idea came up that if the oxides were responsible for the falling emission, a clean metal surface would not cause falling emission when bombarded.

In order to test this hypothesis, a tube was made with a copper strip on the anode. Instead of oxidizing this strip on exhaust, however, hydrogen was introduced and all parts of the anode were heated at 750°C for three minutes in 500 microns pressure of hydrogen. The gas was removed and the process repeated. After this treatment, the copper appeared very bright and it was assumed oxides were reduced.

Here it was seen that with pure Cu (hydrogen fired at exhaust) there is no falling emission.

TABLE I.

Tube No.	Metallic oxide	Filament voltage	Potential applied for falling emission E_a	Retarding cut-off potential E_r	Potential for falling emission E_f
101	FeO	3.5 v	7.0 v	1.5 v	5.5 v
101	FeO	3.1	7.0	1.7	5.3
101	FeO	4.1	6.5	1.3	5.2
103	FeO	4.0	7.0	1.2	5.8
103	FeO	3.7	7.0	1.4	5.6
110	FeO	3.6	7.5	1.95	5.55
110	FeO	4.0	8.0	1.8	6.2
104	CuO	3.4	4.5	2.3	2.2
104	CuO	4.0	4.5	2.4	2.3
105	CuO	3.5	3.5	1.9	1.6
105	CuO	4.1	3.5	1.7	1.7
106	CuO	3.3	4.0	1.9	2.1
106	CuO	3.5	4.0	1.9	2.1
114	MoO ₃	4.1	9.0	1.5	7.5
114	MoO ₃	3.4	9.0	2.4	6.6
115	MoO ₃	4.0	9.5	2.2	7.3
115	MoO ₃	4.0	9.5	2.2	7.3
115	MoO ₃	3.3	9.5	2.4	7.1
116	NiO	3.6	7.0	4.2	2.8
116	NiO	4.2	7.0	4.0	3.0
117	NiO	3.4	4.5	1.8	2.7
117	NiO	4.0	4.5	1.7	2.8
117	NiO	3.2	5.0	2.2	2.8
118	NiO	3.5	6.0	3.8	2.5
118	NiO	4.1	5.5	3.3	2.2
216	WO ₃	3.3	7.0	2.1	4.9
216	WO ₃	3.7	6.5	1.8	4.7
218	WO ₃	4.1	8.0	3.1	4.9
218	WO ₃	3.2	7.5	3.5	4.0
218	WO ₃	4.5	7.5	2.9	4.6
203	Ta ₂ O ₅	3.1	7.5	3.0	4.5
203	Ta ₂ O ₅	3.5	7.0	2.7	4.3
204	Ta ₂ O ₅	3.4	7.0	2.8	5.2
204	Ta ₂ O ₅	3.1	7.5	2.5	5.0
214	Ta ₂ O ₅	3.3	7.5	2.8	4.7
214	Ta ₂ O ₅	3.7	7.5	2.3	5.2
210	ZrO ₂	3.5	4.5	2.6	1.7
210	ZrO ₂	4.1	3.5	2.2	1.3
211	ZrO ₂	4.1	3.5	1.2	1.7
211	ZrO ₂	3.4	4.0	2.1	1.9
212	ZrO ₂	3.6	4.0	2.2	1.8
212	ZrO ₂	3.3	5.0	2.5	2.5
212X	ZrO ₂	3.3	3.5	1.4	2.1
212X	ZrO ₂	3.5	3.5	2.0	1.5

If anything, there seems to be a slight tendency towards rising emission. This may be caused by hydrogen still remaining in the copper metal or it may be caused by some other effect. In any case, this shows that falling emission did not occur when the oxide was not present.

Summary of Data

In Table II we have listed the experimental value of the energy required for falling emission to occur. Next to this value is the value of the heat of formation (H) required for the probable chemical reaction to occur. It is assumed that 23 kilogram calories per mole is equivalent to one electron volt per molecule. E_{real} =electron energy at first sign of falling emission and E_{contact} =contact potential between surface studied and barium oxide cathode (Fig. 6).

PART IV

Conclusions and Discussion

From the section containing the data we might first conclude that the phenomenon of falling emission in this study is related to the anode to which the current is being drawn. If falling emission were caused by something occurring in the cathode itself, i.e., electrolysis or caused by residual gases in the tubes, there would be no reason for the change in E_r (energy necessary to show falling emission) from one anode material to another. However, this critical energy of the electron is different for every chemically different surface. Furthermore, in studying different tubes with the same chemically prepared anode surface, we find the same E_r . It was also shown that when a "clean" metal was bombarded, no falling emission occurred. Falling emission did occur,

TABLE II.

Compound	E_{real} volts	Handbook H (Heat of formation for reaction)	E_{contact} volts	Probably dissociated to:
2FeO	$5.6 \pm .2$ A.D.	5.5 volts	$2.7 \pm .3$ A.D.	$2\text{Fe} + \text{O}_2$
CuO	$2.0 \pm .2$	1.5	$3.2 \pm .1$	$\text{Cu} + \frac{1}{2}\text{O}_2$
MoO ₃	$7.15 \pm .03$	7.4	$3.1 \pm .2$	$\text{Mo} + \frac{3}{2}\text{O}_2$
NiO	$2.7 \pm .2$	2.5	$4.4 \pm .2$	$\text{Ni} + \frac{1}{2}\text{O}_2$
WO ₃	$4.7 \pm .15$	5.5	$3.8 \pm .6$	$\text{W} + \frac{3}{2}\text{O}_2$
Ta ₂ O ₅	$4.8 \pm .3$	unknown	$3.5 \pm .2$	$\text{Ta}_2\text{O}_5 + \frac{1}{2}\text{O}_2$
ZrO ₂	$1.9 \pm .3$	unknown	$3.6 \pm .4$	$\text{ZrO} + \frac{1}{2}\text{O}_2$

A.D. = average deviation.

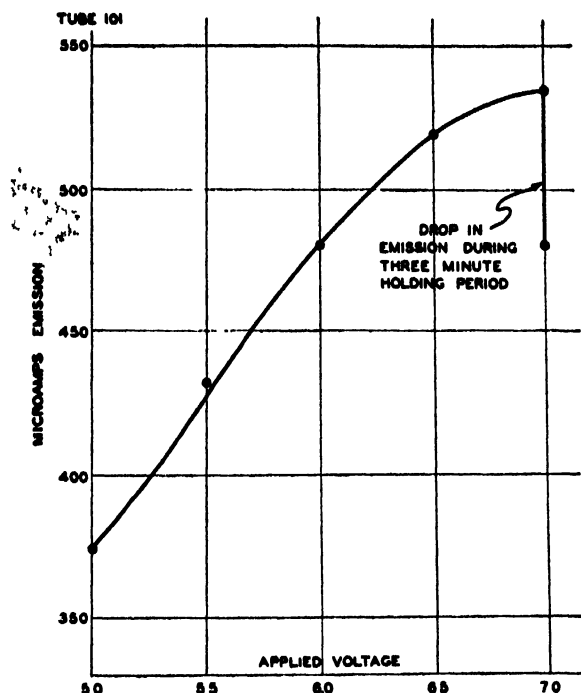


FIG. 5. Falling emission as voltage is increased (three-minute intervals).

however, when this same type of metal (Cu) was oxidized and then bombarded at or above a critical voltage. We might then conclude that the falling emission was a function of some property of the oxide on the anode.

It was seen that a certain amount of energy (electron volts) was needed to cause the effect. When the quantity of energy reaches a critical value, something occurs in the cathode which causes falling emission. We might assume that the bombarded molecule decomposes and that the free oxygen returns to the cathode, combines with the barium on the cathode, and lowers the emission. The statement that oxygen from the oxide is the agent of falling emission is further verified by a relationship in the cases of five metals between the heats of formation and the energy of the electrons at falling emission. These values are found to be equivalent.

There were in general two considerations involved. First, an oxide was chosen which was very common and very likely to be present. Second, the compound in the same series of oxides which had the next lower heat of formation was considered. The difference in energy of

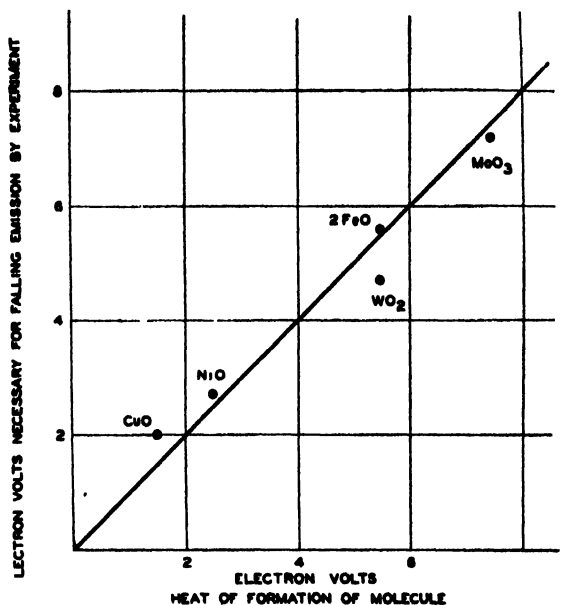


FIG. 6. Relationship between heat of formation of molecules and energy of electron required for falling emission.

the two molecules was taken as the energy required to release oxygen in terms of the heats of formation.

For iron the likelihood of the presence of FeO was high since this compound is common and easy to form. A molecule of 2FeO was assumed to exist. In terms of Heats of Formation approximately 5.5 electron volts would be required to strip this of its oxygen. The value obtained by direct electron bombardment was found to be 5.6 ± 0.2 A.D. electron volts.

In the case of copper, it was assumed that the oxide was CuO. This was done since CuO is one common oxide and Cu_2O is the other. Now the H for Cu_2O is slightly higher than the H for Cu (roughly 39 kg cal. as against 34). Therefore, if energy were supplied to a system, $\text{Cu}_2\text{O} \rightarrow \text{Cu} + \text{CuO}$ and this reaction would not liberate oxygen and thus could not be perceived by measuring emission. However, as more energy was added in the form of higher voltages, the CuO would decompose to $\text{Cu} + \frac{1}{2}\text{O}_2$ and it is this reaction which would affect the emission from the cathode.

In the case of molybdenum, MoO_3 was common and its heat of formation was so close to the experimental value that this compound was chosen. Why the compound seemed to decompose

TABLE III. Work functions of several oxides.

Compound	ϕ of oxide
FeO	3.85 electron volts
CuO	4.35
MoO ₃	4.25
NiO	5.55
WO ₃	4.95
Ta ₂ O ₅	4.65
ZrO ₂	4.75

completely to molybdenum and oxygen instead of going down the transition steps, i.e., to MoO₃ is not known.

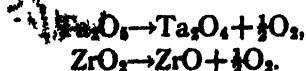
NiO was chosen for nickel since it is the most prominent form and because of good agreement with the H value.

WO₃ was chosen for tungsten since it is a common form of oxide and since the agreement was fairly good with the H listed in the Handbook. It was also noticed in the making of the tube that the yellow WO₃ had not been reached. In one tube not recorded in the data, the oxidizing process was allowed to continue and it took from two to four times the time used to obtain the yellow color.

In considering tantalum and zirconium the same line of reasoning was used. It was assumed that during the oxidation process a common oxide was formed. Since zirconium and tantalum both have a great affinity for oxygen, it was assumed higher oxides were formed, i.e., ZrO₂ and Ta₂O₅.

The H (heat of formation) for both of these compounds is known: the H for their lower oxides, however, have not been reported. The experimental values for the higher forms did not coincide with their values of H ; therefore, just as in the previous cases, it was assumed a low oxide form was reached in the reducing process, rather than free metal.

For tantalum and zirconium the next common lower oxides were considered ZrO and Ta₂O₄. The energy measured thus would be for the following reactions



Using our experimental values of H , we arrived at the conclusion that the lower oxides had the

following values of H :

Ta ₂ O ₅	H —300 kg cal. by Handbook
Ta ₂ O ₄ + $\frac{1}{2}$ O ₂	H —110 kg cal. by measurement
Ta ₂ O ₄	H —190 kg cal. by calculation
ZrO ₂	H —178 kg cal. by Handbook
ZrO + $\frac{1}{2}$ O ₂	H —43 kg cal. by measurement
ZrO	H —135 kg cal. by calculation

PART V

Further Considerations of Conclusions

(1) Conclusion No. 1 concerns the mechanism by which falling emission occurs. It describes a manner in which oxides may decompose when bombarded in vacuum and explains why oxidized parts are in general harmful when used in vacuum tube construction when subjected to electron bombardment. It also suggests the possibility that other compounds may decompose in a similar manner upon bombardment, i.e., salts, silicates, etc. As far as the author knows, no tests of this type have been reported in the literature.

(2) Utilizing the data concerning contact potentials the work functions of various oxide films may be determined.

By ϕ oxide = E contact + 1.15 volts (ϕ' of Ba:BaO).¹¹

Values found experimentally for the contact potentials were used to compile Table III.

(3) We have found in the methods outlined a means of predicting the heats of formation of two compounds the H value of which has not been reported in the literature.

$$\begin{aligned} H \text{ Ta}_2\text{O}_4 &= 190 \text{ kg cal. per mole,} \\ H \text{ ZrO} &= 135 \text{ kg cal. per mole.} \end{aligned}$$

This method may possibly be used further in determining the heats of formation of other compounds whose values are undetermined.

There is one more point worthy of mention. While the voltages found necessary to decompose the oxides correspond to the heats of formation, it may not be the actual heats of formation that are being studied, but rather the free energy of dissociation. In this case however, the heats of

¹¹ J. H. DeBoer, reference 7, p. 357. This value of work function is not identical with all experiments. Thus there is an error probably of ± 3 volt in this value and the values of oxides listed below.

formation and the free energies of dissociation are so close that the means by which we measure these quantities would not separate them. Actually then, we may be reading free energies of dissociation by our measurements rather than heats of formation.

Lewis and Randall give the following relationship¹³

$$F = H - TS = (E + PV) - TS,$$

where

F —free energy per mole,
 H —heat of formation per mole,
 T —temperature,
 E —internal energy,
 P —pressure,
 V —volume, and
 S —entropy.

It is known that there is a slight variation in F and H with P and T , but in our case $P = 0$ by

¹³ G. N. Lewis and M. Randall, *Thermodynamics* (McGraw-Hill Book Company, Inc., New York, 1923).

approximation, S is small for metallic oxides,¹⁴ even though the temperature of the anode oxides is slightly higher than room temperature. Hence, F is nearly equal to H which is nearly equal to E , and we have a rough means of measuring these quantities in the method described above.

- (a) Free energy of dissociation per mole, F ;
- (b) Heat of formation per mole, H ;
- (c) Internal energy per mole, E .

Acknowledgment should be made to Dr. E. A. Lederer for suggesting this problem and to Mr. D. H. Wamsley, Dr. L. B. Headrick, Mr. G. Widell, and Dr. L. Malter of the RCA Laboratories and Professors S. A. Korff, C. W. Barnes, and A. F. Myers of New York University for their suggestions pertaining to this work.

¹⁴ Bulletin 384, U. S. Department of the Interior, Bureau of Mines (Entropy of CaO is reported to be 9.5 calories at 298.1° absolute, page 16. Entropy for CuO—10.4 calories at same temperature, page 19. Entropy for FeO—14.2 calories at same temperature, page 20).

An Experimental Investigation of Forced Vibrations in a Mechanical System Having a Non-Linear Restoring Force

CARL A. LUDEKE

Department of Mathematics and Mechanics, University of Cincinnati, Cincinnati, Ohio

(Received April 4, 1946)

This paper introduces a mechanical apparatus capable of generating and recording forced vibrations in a system having a non-linear restoring force. The experimental wave forms are then compared with the theoretical results given by three graphical methods. These are the methods of Martienssen, Hartog, and Rauscher. Provided a suitable first guess for the amplitude can be made, one application of Rauscher's method gives better results than the other methods. A graph is made of the experimental amplitude as a function of the disturbing frequency. These results are compared with the theoretical results of the above three methods. Again one application of Rauscher's method gives satisfactory results. In all experimental results it is noted that even though the restoring force is distinctly non-linear, the wave forms of the resulting motion are nearly sinusoidal as long as the frequency of the observed motion is the same as the frequency of the disturbing force. However, steady oscillations can be maintained for which the observed frequency is a sub-multiple of the disturbing frequency. Two such subharmonics were recorded and the experimental wave forms are shown.

CONSIDER an undamped system of one degree of freedom with a curved spring characteristic and under the influence of a harmonic disturbing force. The equation of motion is given by:

$$m(d^2x/dt^2) + f(x) = P_0 \cos \omega t, \quad (1)$$

in which m is the mass of the system, $f(x)$ represents the spring characteristic, P_0 is the amplitude of the disturbing force, and ω is the angular frequency of the disturbing force. An exact solution of Eq. (1) exists only for some very special forms of $f(x)$. There are however three approximate methods available, namely:

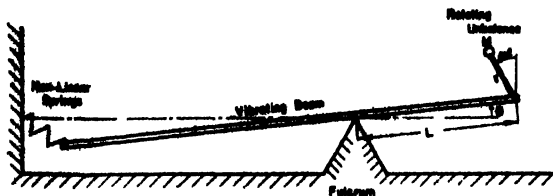


FIG. 1. Diagram of system.

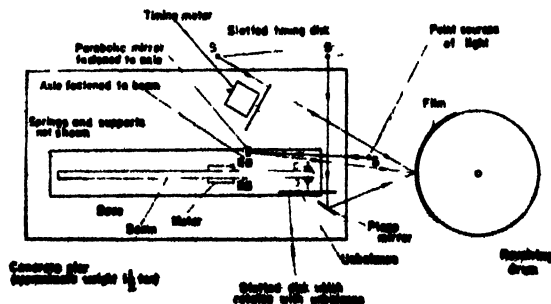


FIG. 2. Diagram of experimental arrangement.

the methods of Martienssen,¹ Hartog,² and Rauscher.³ In each of these methods $f(x)$ need only be known graphically, and the methods themselves can then be completed by an entirely graphical procedure. The method of Martienssen gives a one-term solution, the method of Hartog gives a two-term solution, and the method of Rauscher gives a graphical solution capable of being repeated until the desired accuracy is obtained. In this paper only one application of the Rauscher method is used. It would be of interest therefore to compare the results of the three approximate methods with exact solutions if they were in general available. Since this is not the case, they are compared with experimental results.

These results were obtained by means of the system shown in the diagram of Fig. 1. By changing springs a variety of load deflection curves could be obtained; all however of the "increasing stiffness" type. Every effort was made to keep the damping forces to a minimum. The unbalance was rotated with a constant angular velocity by means of an electric motor.

¹ O. Martienssen, *Physik. Zeits.* 11, 448-460 (1910).

² J. P. Den Hartog, "The amplitudes of non-harmonic vibrations," *J. Frank. Inst.* 216, No. 4, 459-473 (October, 1933).

³ M. Rauscher, "Steady oscillations of systems with non-linear and unsymmetrical elasticity," *J. App. Mech.* 8, A169-A177 (1938).

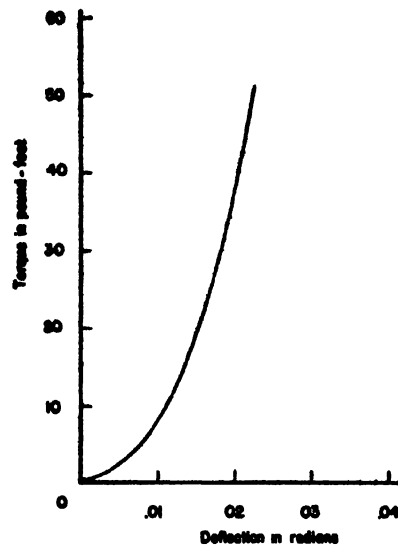


FIG. 3. Load deflection curve.

The equation of motion for this system is then

$$I(d^2\theta/dt^2) + f(\theta) = MrL\omega^2 \cos \omega t = P_0 \cos \omega t, \quad (2)$$

in which

I is the moment of inertia of the vibrating parts = .31 slug-ft.²,

$f(\theta)$ is the spring torque,

ω is the constant angular velocity of the unbalance,

M is the mass of the unbalance,

r is the radius of the unbalance,

L is the distance from the center of rotation of the unbalance to the center of rotation of the beam, and

$$P_0 = MrL\omega^2.$$

A diagram of the experimental set-up is shown in Fig. 2. A typical load deflection curve is

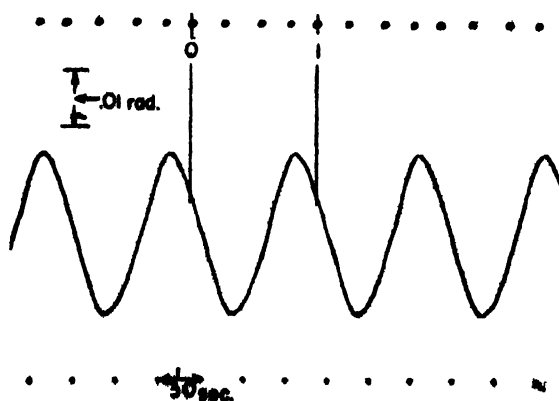


FIG. 4. Recorded wave form when using the load deflection curve of Fig. 3. $\omega = 64.8$ cycles/sec.; $P_0 = .372$ ft.-lb.

shown in Fig. 3. This spring characteristic was used when recording the wave form shown in Fig. 4. The series of dots at the bottom of the picture represent time intervals of 1/30 sec. The series of dots at the top represent the position of the unbalance at 90° intervals. The heavier dot indicates that the unbalance is vertically upward, i.e., that $\omega t = 0$. The simplest approximate solution¹ of Eq. (1) is based on the assumption that the motion $x = f(t)$ is sinusoidal and has the "forced" frequency. That is, the solution is assumed to be

$$x = x_0 \cos \omega t. \quad (3)$$

Using this solution, the forces involved are the inertia force $-m\omega^2 x_0 \cos \omega t$, the external force $P_0 \cos \omega t$, and the restoring force $f(x)$. We note that these are all zero when $x = 0$, i.e., Eq. (1) is satisfied by Eq. (3) when $x = 0$. We shall now require the solution to satisfy Eq. (1) when $x = x_0$. Thus

$$-m\omega^2 x_0 + f(x_0) = P_0$$

or

$$f(x_0) = m\omega^2 x_0 + P_0. \quad (4)$$

A very convenient way of finding the value of x_0 which satisfies Eq. (4) is as follows: Plot $f(x)$ load *versus* deflection, superimpose on this curve a straight line having a load intercept P_0 and a slope $m\omega^2$. The point of intersection of the straight line and the curve will determine x_0 . This value is then used in Eq. (3) to give the approximate solution. Obviously this method is only true for all values of t when $f(x)$ is linear, for any other form of $f(x)$ it is true only at the middle and the end of the "stroke."

Figure 5 shows a load deflection curve with the construction lines for the one-term solutions. The experimentally recorded wave forms are shown in Fig. 6, together with the Martienssen solutions. The Hartog solution uses the same basic idea of the Martienssen solution, namely, to hold the frequency fixed and to choose an amplitude corresponding to it; but it extends the solution to two terms instead of one term. This allows us to balance the forces at three points of the "stroke" instead of at only two points of the stroke.

In this method, the solution of Eq. (1) is assumed to be of the form

$$x = a_1 \cos \omega t + a_2 \cos 3\omega t, \quad (5)$$

and $f(x)$ is assumed to be of the form

$$f(x) = A_1 \cos \omega t + A_2 \cos 3\omega t. \quad (6)$$

Substituting Eqs. (5) and (6) in Eq. (1) we have

$$(A_1 - m\omega^2 a_1 - P_0) \cos \omega t + (A_2 - 9m\omega^2 a_2) \cos 3\omega t = 0. \quad (7)$$

Equation (7) must be satisfied for all values of t . Thus

$$A_1 = m\omega^2 a_1 + P_0, \quad A_2 = 9m\omega^2 a_2. \quad (8)$$

Therefore Eq. (6) takes the form

$$f(x) = (m\omega^2 a_1 + P_0) \cos \omega t + 9m\omega^2 a_2 \cos 3\omega t. \quad (9)$$

We must now choose the a 's in Eq. (5) so that when the A 's are calculated by means of Eq. (8), $f(x)$ as represented by Eq. (9) turns out correctly.

We notice that when $\omega t = \pi/2$, Eq. (7) is satisfied whether or not Eqs. (8) are satisfied. Since Eqs. (8) contain two parameters a_1 and a_2 , we can satisfy Eq. (7) at two other values of ωt . Let us choose $\omega t = 0$, and $\omega t = \pi/6$.

At $\omega t = 0$, $x = a_1 + a_2$;

$$f(x) = m\omega^2 a_1 + P_0 + 9m\omega^2 a_2 \quad (10)$$

at $\omega t = \pi/6$, $x = \sqrt{3}a_1/2$;

$$f(x) = (m\omega^2 a_1 + P_0)\sqrt{3}/2.$$

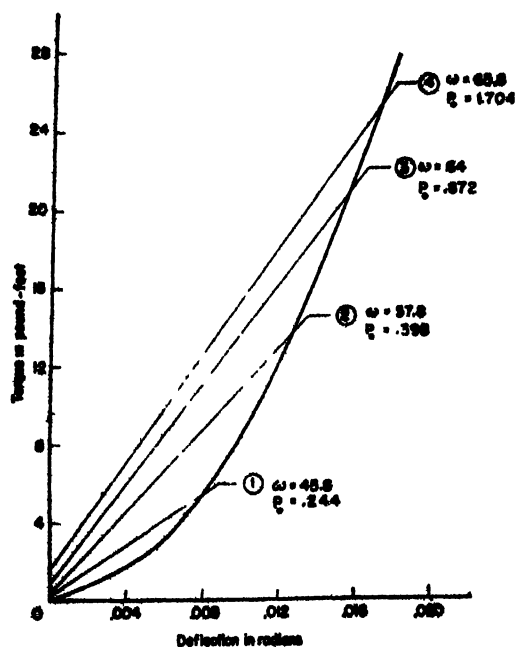


FIG. 5. Load deflection curve with construction lines for one-term solution.

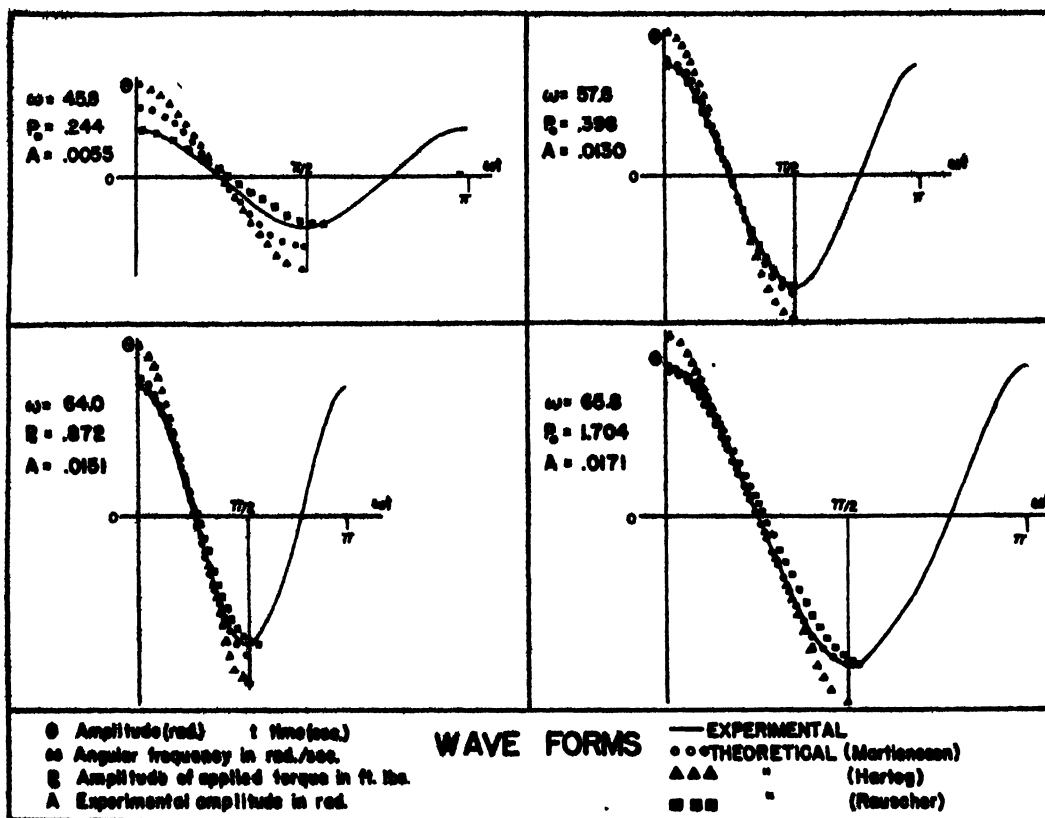


FIG. 6. Wave form recorded when using the load deflection curve of Fig. 5.

These relations can be written in the form

$$f\left(\frac{a_1\sqrt{3}}{2}\right) = m\omega^2\left(\frac{a_1\sqrt{3}}{2}\right) + P_0\frac{\sqrt{3}}{2}, \quad (11)$$

$$f(a_1 + a_2) = 9m\omega^2(a_1 + a_2) + (P_0 - 8m\omega^2 a_1).$$

The right sides of Eqs. (11) represent straight lines with slopes $m\omega^2$ and $9m\omega^2$, respectively, and having intercepts $P_0\sqrt{3}/2$ and $(P_0 - 8m\omega^2 a_1)$, respectively. The left sides of Eqs. (11) represent the function $f(x)$ evaluated at $x = a_1\sqrt{3}/2$ and $x = a_1 + a_2$. Thus a_1 and a_2 can be solved for graphically as shown in Fig. 7. The results for the load deflection curve of Fig. 5 are shown on Fig. 6.

In both methods presented thus far the procedure was to determine an amplitude corresponding to a given frequency. Perhaps the most exact method available does just the opposite. It chooses an amplitude and then solves for the corresponding frequency. It is a method which can be repeated until the desired accuracy is

attained. The main problem is to make a suitable first choice for the amplitude.

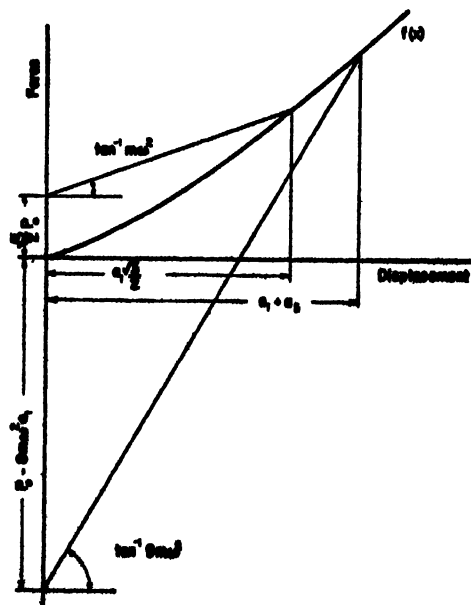


FIG. 7. The graphical method of finding a_1 and a_2 . Two-term solution.

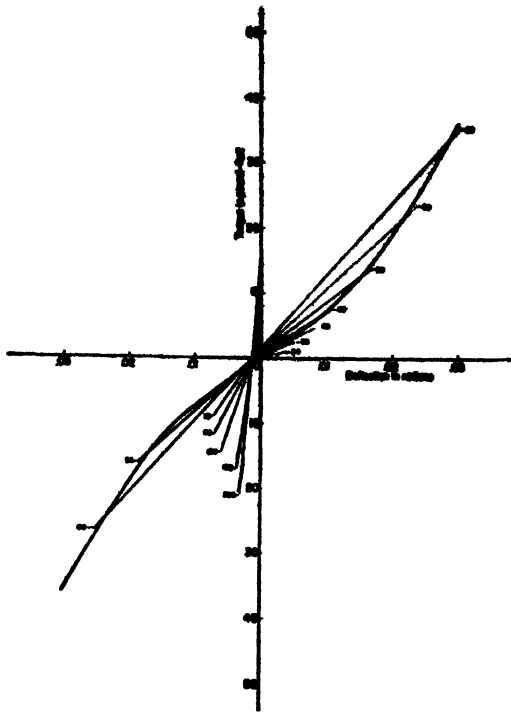


FIG. 8. Load deflection curve.

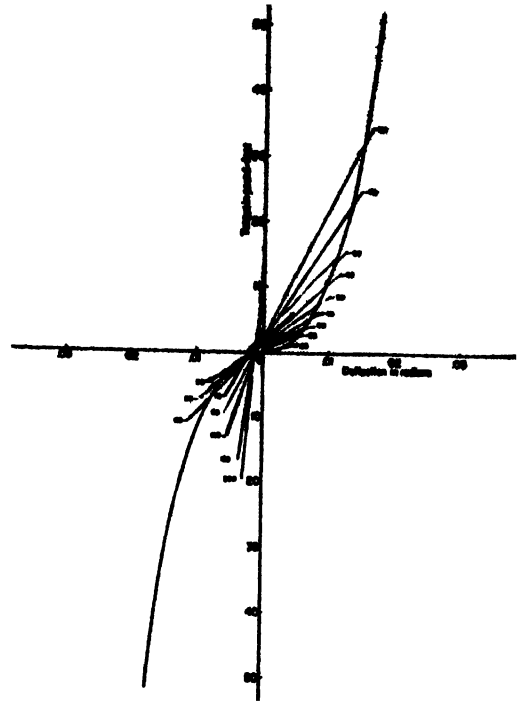


FIG. 10. Load deflection curve.

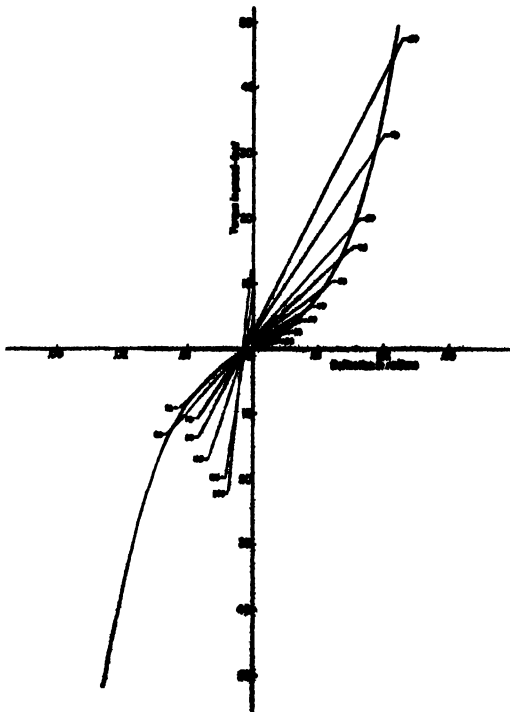


FIG. 9. Load deflection curve.

In this method Eq. (1) is reduced to an equation in which the right side is zero, i.e.,

$$m(d^2x/dt^2) + R(x, t) = 0, \quad (12)$$

in which $R(x, t) = f(x) - P_0 \cos \omega t$. In a system in which $f(x)$ is linear, the solution is $x = A \cos \omega t$, hence R can be written as a function of x alone, i.e.,

$$R(x) = f(x) - P_0 x/A. \quad (13)$$

For systems which are slightly non-linear this equation can be assumed valid. Thus Eq. (1) takes the form

$$m(d^2x/dt^2) + R(x) = 0, \quad (14)$$

in which R is given by Eq. (13).

This equation of a free vibration is solved as follows: A first integral of Eq. (14) gives us:

$$v^2 = \frac{2}{m} \int_a^A R(x) dx; \quad v = \frac{dx}{dt}. \quad (15)$$

The time corresponding to any displacement x is then found from the equation

$$t = \int_a^A \frac{dt}{dx} dx = \int_a^A \frac{dx}{v}. \quad (16)$$

For a purely graphical evaluation of this integral a change of variable is introduced by letting

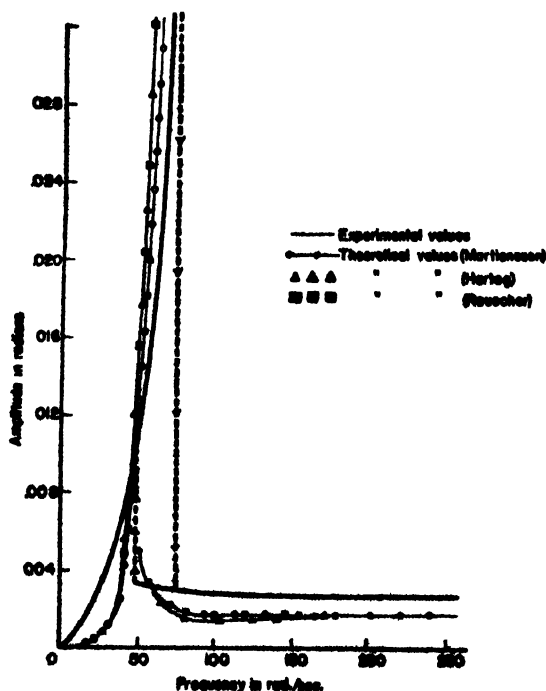


FIG. 11. Amplitude versus frequency when using load deflection curve of Fig. 8.

$x = A \sin \varphi$. Thus Eq. (16) becomes

$$t = \int_0^{\pi/2} \frac{A \cos \varphi}{v} d\varphi. \quad (17)$$

The procedure is therefore as follows: After choosing A , a table of values of x and v is made by graphically integrating Eq. (15). The variable x is changed to φ by means of the relation $x = A \sin \varphi$, and a new table of values of φ and t is made by graphically integrating Eq. (17). From these two tables a plot is made of x and t , and the frequency corresponding to the choice of amplitude A is then determined. The entire process can then be repeated by using the relationship between x and t to determine a more exact value of the function $R(x)$ used in Eq. (15).

Having the experimental results as shown in Fig. (6), the initial choice of A is greatly simplified. Choosing A equal to the experimental values of the amplitude, the results of one application of this method are shown on Fig. 6.

In Figs. 8-10 are shown three load deflection curves of varying non-linearity. On these curves are superimposed the construction lines for one term solutions. From these results we can plot a

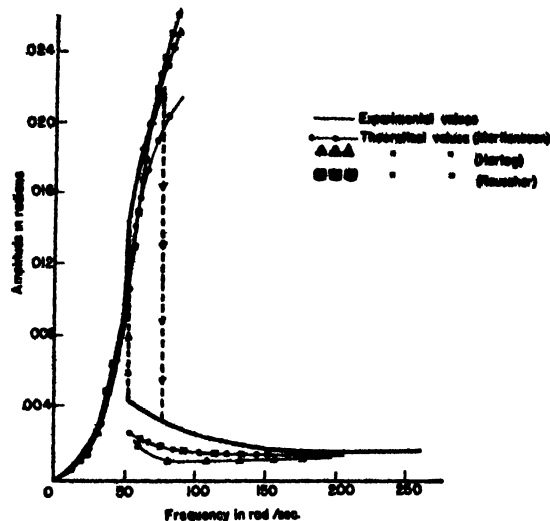


FIG. 12. Amplitude versus frequency when using load deflection curve of Fig. 9.

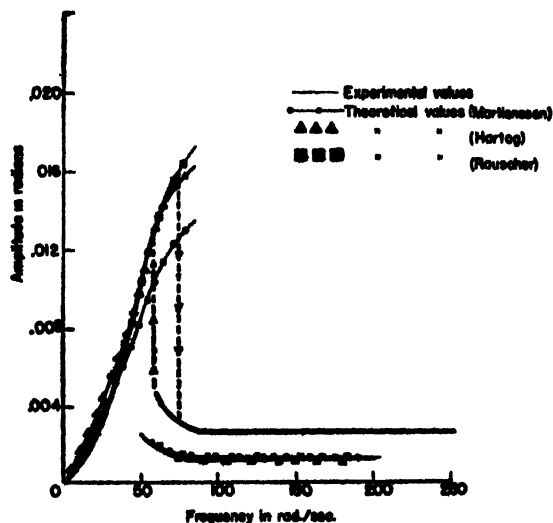


FIG. 13. Amplitude versus frequency when using load deflection curve of Fig. 10.

relationship between amplitude and frequency. The experimental results together with the theoretical results of all three methods are shown in Figs. 11-13. The lack of agreement between the experimental results and theoretical results for the small amplitudes observed at frequencies higher than the "jump" frequency is caused primarily by the errors involved in recording the small amplitudes and in determining accurately enough the load deflection curves for very small amplitudes.

In the three theoretical methods illustrated,

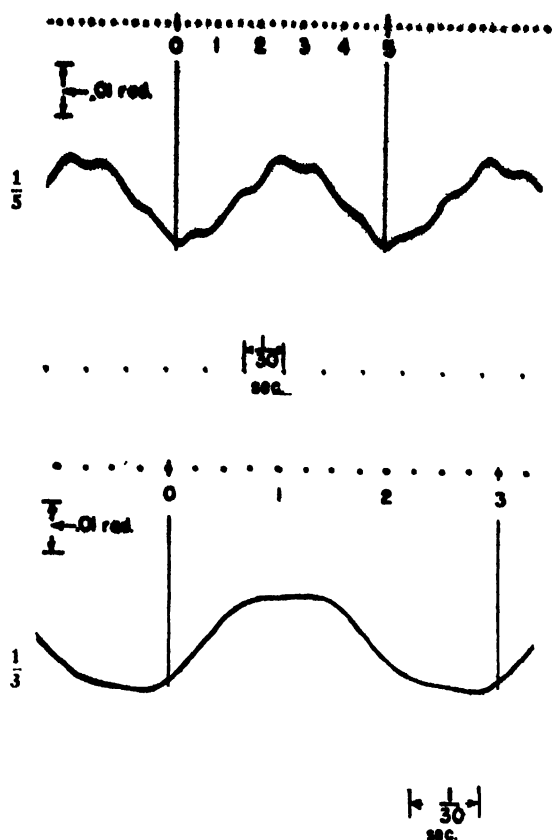


FIG. 14. Sub-harmonic resonance wave forms recorded when using the load deflection curve of Fig. 15.

the initial assumption was made that the solution of Eq. (1) should have a frequency approximately equal to the frequency of the disturbing force. This, however, does not cover all possible solutions. In fact one of the most interesting phenomena observed in non-linear vibration systems is that of sub-harmonic resonance.⁴ This is the phenomena of a steady forced vibration whose frequency is a sub-multiple of the disturbing frequency. The recorded wave forms of sub-harmonic resonance ($\frac{1}{3}$) and ($\frac{1}{2}$) are shown in Fig. 14. These were observed when using the load deflection curve shown in Fig. 15. The three theoretical methods as used in this paper would not apply to these wave forms.

Before drawing final conclusions, let us empha-

⁴ C. A. Ludeke, "Resonance," J. App. Phys. 13, No. 7, 410-423 (July 1942).

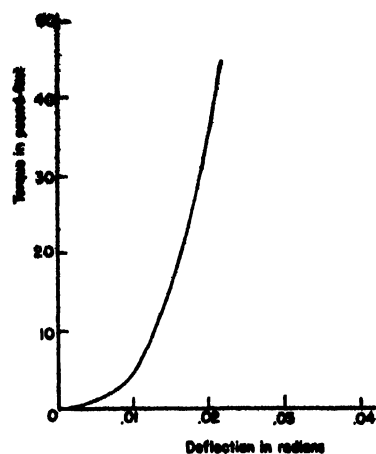


FIG. 15. Load deflection curve.

size the fact that the theoretical methods were applied as if the system contained no damping. This is of course impossible to achieve experimentally. Also, there is considerable experimental error when observing frequencies higher than the jump frequency. Bearing this in mind, however, we can draw the following conclusions. First, even though the load deflection curve varies considerably from a straight line, as indicated in Fig. 3, and even though the amplitude of the maintained vibration is quite large, as indicated in Fig. 6, $A=0.0171$, nevertheless the resulting wave forms, Fig. 6, are very nearly sinusoidal. This means that the very simple approximation of Martienssen is nevertheless a very good one, at least for the cases herein considered. Second, though the two-term method at some particular frequency gives the exact amplitude, see Figs. 11-13, nevertheless the values depend so critically on the values of P_0 and ω that for the most part the results are not as good as the one-term solution, see Fig. 6. Third, provided the experimental amplitude is known, the short cut method of Rauscher can be applied once with excellent results. Fourth, as soon as the case of subharmonic resonance is considered the recorded wave forms are no longer sinusoidal.

It is a pleasure to thank Miss M. Seuberling for her assistance in preparing the drawings.

A New Method for Measuring Dielectric Constant and Loss in the Range of Centimeter Waves*

S. ROBERTS AND A. VON HIPPEL

Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received April 29, 1946)

In 1940, dielectric measurements in the centimeter range were considered as difficult and not very accurate. The authors, therefore, developed a "hollow-pipe" method which overcame these objections and required only a weak oscillator and small amounts of the dielectric material. The theory and its practical applications, as perfected by March 1941, are presented in this paper.

THE capacitance of a condenser is proportional to the dielectric constant or permittivity ϵ^* of the dielectric material which it contains.

$$C = (\epsilon^* C_0) / \epsilon_0, \quad (1)$$

where C_0 is called the vacuum capacitance and ϵ_0 is the permittivity of free space. When subjected to a sinusoidal voltage V , a condenser passes a current I which normally does not lead the voltage by 90° but only by $(90^\circ - \delta)$ (Fig. 1). If the current is defined in the general way

$$I = j\omega CV = j\omega C_0 \frac{\epsilon^*}{\epsilon_0} V, \quad (2)$$

the existence of the loss angle δ can be represented by introducing a complex permittivity

$$\epsilon^* = \epsilon' - j\epsilon''. \quad (3)$$

The charging current is proportional to ϵ' , the loss current to ϵ'' , and the loss tangent

$$\tan \delta = \epsilon'' / \epsilon' \quad (4)$$

is given by the ratio of loss current to charging current. ϵ' and ϵ'' or ϵ' and $\tan \delta$ are the parameters normally used for describing the frequency response of a dielectric material. The loss tangent is identical with the power factor $\cos \theta$ for small angles δ .

* This research, about which a preliminary report was given at the meeting of the American Physical Society in April, 1940, was ready for publication by March, 1941, but had to be classified and withheld from general circulation. The method, now one of common knowledge, has been refined further in our laboratory during the war and is being used with coaxial lines and hollow wave guides in the range from one meter to one centimeter. A survey of this method, development and of the extensive dielectric information obtained will be given elsewhere.

The "lumped" circuit aspect used above becomes invalid in the ultra-high frequency range, but the concept of a complex dielectric constant is not affected. It describes the physical fact that a dielectric in a periodic electromagnetic field carries an electric current of the density

$$J = \sigma E + \partial D / \partial t = (\epsilon' - j\epsilon'') j\omega E, \quad (5)$$

consisting of two components, a charging current of the density

$$J_c = \epsilon' j\omega E \quad (6a)$$

storing electric energy by the reversible displacement of charge carriers, and a loss component

$$J_l = \epsilon'' \omega E \quad (6b)$$

measuring the dissipation of energy in heat caused by traveling charge carriers as well as by the time lag between the electric field E and the dielectric flux density D .

The dielectric constant and loss can be found indirectly in the microwave range by transmission line or optical methods. In the former ϵ^* is derived from the intrinsic wave impedance Z_0 or propagation function γ_0 of the dielectric medium, while in optical methods it is evaluated from the complex index of refraction n^* equal to $n(1 - j\kappa)$. Standing or traveling waves can serve for meas-

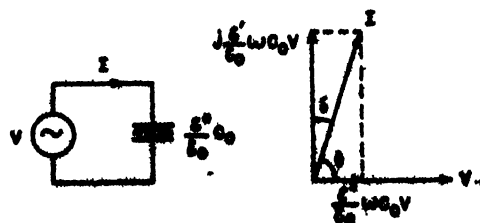


FIG. 1. Relation between current and voltage in a condenser with loss.

using any one of these interrelated sets of parameters.

The normal optical procedure employs traveling waves. The intensities of the reflected and transmitted beam are measured and n and k evaluated by Fresnel's equations. Esau, with his co-workers Báz¹ and Kebbel,² has carried this method over into the ultra-high frequency range, but here it becomes definitely inferior. Serious boundary problems arise because the wave-length is of the dimensions of the sample, and standing waves are set up between transmitter and receiver, producing intensity fluctuations by interference.

The macroscopical wave-length, on the other hand, proves of great advantage if standing waves, that is interference optics, are applied because a detector may travel directly through the profile of the wave pattern. Drude's two classical methods³ utilize this possibility and have since been employed in many variations.⁴ However, a handicap of the Lecher system remained: an empirical calibration of the condenser system was necessary, and extreme care had to be taken to avoid perturbation of the waves by the detector system and to shield the detector against stray fields. These limitations have been overcome by the development of a "hollow-pipe" method, to be described in the following paragraphs.

PRINCIPLE OF THE METHOD

By limiting the electromagnetic field to the enclosure of a hollow pipe or coaxial line, all boundary and stray effects disappear automatically and small amounts of any dielectric can be measured with precision. With this concept in mind we developed the apparatus shown schematically in Fig. 2. A transmitter radiates waves of a given frequency into one end of a closed wave guide; they are reflected by the metallic boundary at the other end. Standing waves are set up and can be measured by a probe detector traveling along a slot in the pipe parallel

to its axis. The dielectric is inserted in the closed end of the pipe opposite to the transmitter, filling the volume to a height d . Above it the wave pattern is measured in air. ϵ' and ϵ'' are calculated from the ratio of the field strength in node and

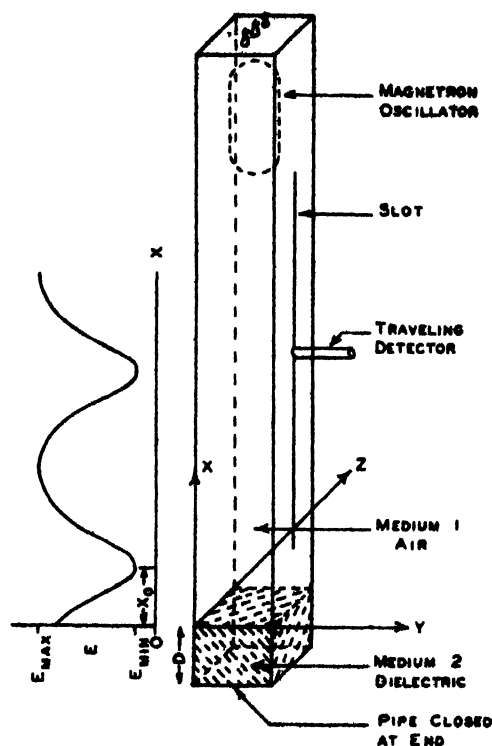


FIG. 2. Diagram of the apparatus showing principle of the method.

antinode $E_{\min}/E_{\max}$ and the distance x_0 of the first node from the surface of the dielectric.

It is a known procedure to calculate the terminating impedance of a transmission line from the voltage ratio $V_{\min}/V_{\max}$ and the distance x_0 ,⁵ or an acoustic impedance, by the location and relative magnitude of the sound-pressure maxima and minima.⁶ But measurements of the terminal impedance of a short wave guide filled with a dielectric material have not been applied, so far as we know, to the evaluation of ϵ' and ϵ'' . We are giving, therefore, in the following pages a mathematical theory of the method, the description of our apparatus, and some results illustrating its performance.

⁵ H. O. Rosenstein, Zeits. f. Hochfrequenztechnik 36, 81 (1930); N. N. Malov, Physik. Zeits. Sowjetunion 11, 539 (1937).

⁶ W. M. Hall, J. Acous. Soc. Am. 11, 141 (1939).

¹ G. Báz, Physik. Zeits. 40, 394 (1939).

² W. Kebbel, Hochtech. u. Elek. Akus. 53, 81 (1939).

³ P. Drude, Wied. Ann. 55, 633 (1895); 61, 466 (1897).

⁴ See, for instance, G. Potapenko, Zeits. f. Physik 20, 26 (1923); M. Seelinger, Ann. d. Physik 16, 77 (1933); R. King, Rev. Sci. Inst. 8, 201 (1937); K. Stewart, Ann. d. Physik 36, 141 (1939); H. Silitis, Ann. d. Physik 36, 397 (1939).

MATHEMATICAL THEORY

A standing wave in the pipe can be represented as the sum of two traveling waves progressing in opposite directions parallel to the axis of symmetry. The transverse field components $E(x)_1$ and $H(x)_1$ in air (medium 1) at a variable distance x from the dielectric surface along a line parallel to the axis are given below.

$$\begin{aligned} E(x)_1 &= A_{11} \exp(\gamma_1 x) + A_{r1} \exp(-\gamma_1 x) \\ &= A_{11} [\exp(\gamma_1 x) + r_0 \exp(-\gamma_1 x)], \end{aligned} \quad (7)$$

$$\begin{aligned} H(x)_1 &= \frac{A_{11}}{Z_1} \exp(\gamma_1 x) - \frac{A_{r1}}{Z_1} \exp(-\gamma_1 x) \\ &= \frac{A_{11}}{Z_1} [\exp(\gamma_1 x) - r_0 \exp(-\gamma_1 x)]. \end{aligned}$$

A_{11} and A_{r1} are the amplitudes of the incident and reflected waves at the dielectric surface. Their ratio

$$A_{r1}/A_{11} = r_0 = e^{-2\phi}, \quad (8)$$

where

$$\phi = \rho + j\psi,$$

defines a reflection coefficient which characterizes the terminating pipe section containing the dielectric medium 2. The wave impedance $Z(0)$ of this section, given by the ratio of the electric to the magnetic field intensities at the dielectric surface, is related to r_0 and ϕ .

$$Z(0) = \frac{E(0)}{H(0)} = Z_1 \frac{1+r_0}{1-r_0} = Z_1 \frac{1+e^{-2\phi}}{1-e^{-2\phi}} = Z_1 \coth \phi. \quad (9)$$

Normally the attenuation in the air-filled part of the pipe can be neglected, that is, the propagation γ_1 , is reduced to its phase factor

$$\gamma_1 = j\beta_1 = j2\pi/\lambda_1, \quad (10)$$

where λ_1 is the wave-length in air inside the waveguide. Since γ_1 is imaginary, that is, $|\exp(\pm\gamma_1 x)| = 1$, it follows from (7) that the maximum and minimum field amplitudes of the standing wave are

$$\begin{aligned} E_{\max} &= |A_{11}|(1+|r_0|) = |A_{11}|(1+e^{-2\rho}), \\ E_{\min} &= |A_{11}|(1-|r_0|) = |A_{11}|(1-e^{-2\rho}), \end{aligned} \quad (11)$$

and their ratio is

$$\frac{E_{\min}}{E_{\max}} = \frac{1-e^{-2\rho}}{1+e^{-2\rho}} = \tanh \rho. \quad (12)$$

The first minimum above the dielectric boundary occurs at a distance x_0 where the incident and reflected electric waves have a phase difference of π radians, that is,

$$\frac{2\pi x_0}{\lambda_1} = -2\psi - \frac{2\pi x_0}{\lambda_1} - \pi, \quad (13a)$$

or

$$\psi = 2\pi(\frac{1}{2} - x_0/\lambda_1). \quad (13b)$$

The wave impedance $Z(0)$ can now be expressed in terms of the directly measurable parameters x_0 , λ_1 , and $E_{\min}/E_{\max}$. Expanding $\coth \phi$ in Eq. (9) one obtains

$$\begin{aligned} Z(0) &= Z_1 \frac{\tanh \rho - j \cot \psi}{1 - j \tanh \rho \cot \psi} \\ &= Z_1 \frac{\frac{E_{\min}}{E_{\max}} - j \tan \frac{2\pi x_0}{\lambda_1}}{1 - j \frac{E_{\min}}{E_{\max}} \tan \frac{2\pi x_0}{\lambda_1}}. \end{aligned} \quad (14)$$

The impedance $Z(0)$ is determined by the length d and the propagation function γ_2 of the short section of pipe containing the dielectric, or conversely, if $Z(0)$ and d are known, γ_2 can be derived from them.

The standing wave in the dielectric is

$$E(x)_2 = A_{12} \exp(\gamma_2 x) + A_{r2} \exp(-\gamma_2 x), \quad (15)$$

$$H(x)_2 = \frac{A_{12}}{Z_2} \exp(\gamma_2 x) - \frac{A_{r2}}{Z_2} \exp(-\gamma_2 x),$$

and at $x = -d$ (Fig. 2) the pipe is terminated by a metal plate, introducing at this boundary a voltage node.

$$0 = A_{12} \exp(-\gamma_2 d) + A_{r2} \exp(\gamma_2 d). \quad (16)$$

Solving (16) for A_{r2}/A_{12} and substituting this value in (15) gives

$$Z(0) = \frac{E(0)_2}{H(0)_2} = Z_2 \tanh \gamma_2 d. \quad (17)$$

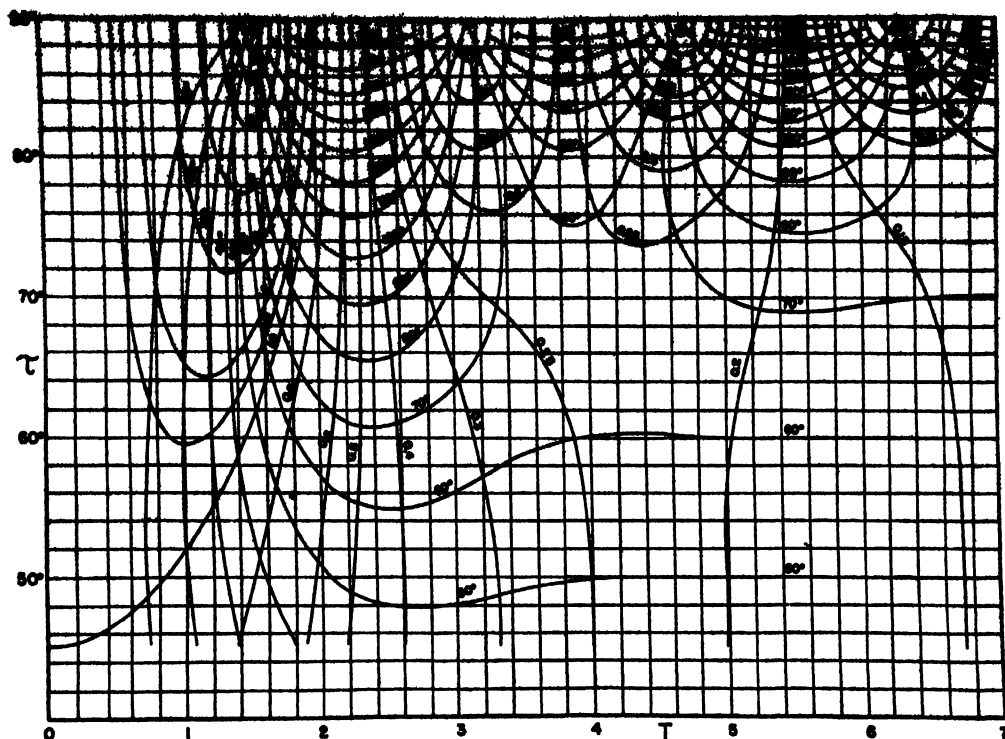


FIG. 3. Chart of the function $Ce^{it} = \tanh Te^{i\tau} / Te^{i\tau}$, with ζ and τ expressed in degrees.

The characteristic impedance Z_2 and the propagation function γ_2 of any TEM or TE wave in a dielectric are related in the following manner:

$$\gamma_2 Z_2 = j\omega\mu_2, \quad (18a)$$

where μ_2 is the permeability of the dielectric medium. The permeability of dielectric materials is generally μ_0 of free space; therefore

$$\gamma_2 Z_2 = j\omega\mu_0 = \gamma_1 Z_1, \quad (18b)$$

and

$$Z(0) = Z_1 \frac{\gamma_1}{\gamma_2} \tanh \gamma_2 d. \quad (19)$$

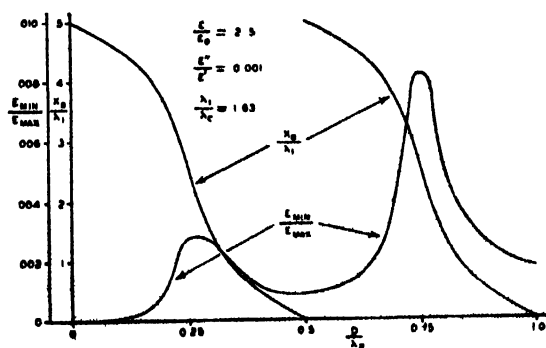


FIG. 4. $E_{\min}/E_{\max}$ and x_0/λ_1 , versus the ratio of thickness d of the sample to wave-length λ_1 in the dielectric.

Equating (19) and (14) yields

$$\frac{\tanh \gamma_2 d}{\gamma_2 d} = \frac{-j\lambda_1 \frac{E_{\min}}{E_{\max}} - j \tan \frac{2\pi x_0}{\lambda_1}}{2\pi d \frac{1 - j \frac{E_{\min}}{E_{\max}} \tan \frac{2\pi x_0}{\lambda_1}}}{= Ce^{it}. \quad (20)$$

The function Ce^{it} can be found, as shown above, by measuring the thickness of the sample d , the wave-length λ_1 in the air-filled pipe, the ratio of minimum to maximum field strength $E_{\min}/E_{\max}$, and the distance x_0 of the first minimum from the dielectric surface.

The next step is the determination of $Te^{i\tau} = \gamma_2 d$ from charts or from a series approximation of the function

$$\frac{\tanh Te^{i\tau}}{Te^{i\tau}} = Ce^{it}. \quad (21)$$

In Fig. 3 the argument τ is the ordinate, the absolute value T the abscissa, while C and ζ are parameters of intersecting curves. τ and ζ are expressed in degrees. The hyperbolic functions are multivalued so that measurements with a

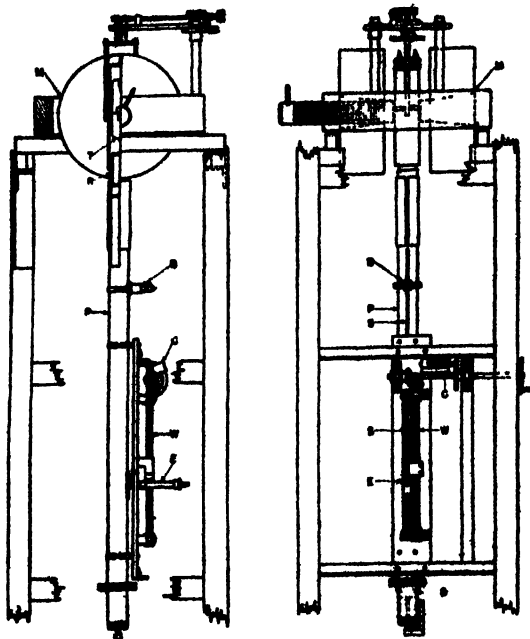


FIG. 5. Constructional details of the apparatus.

single thickness d of the dielectric may leave the value of γ_2 in doubt. But if two different thicknesses are used, only one set of values T/d and τ , will satisfy both experimental results.

For low loss materials it is desirable to use a thickness near $\lambda_2/4$ or higher odd multiples of $\lambda_2/4$. As Fig. 4 shows, $E_{\min}/E_{\max}$ reaches maxima for these values of thickness. The apparatus used in 1940 could measure the loss if $E_{\min}/E_{\max}$ were larger than 0.003. For $d=\lambda_2/4$ in Fig. 4, the loss would be below this limit, while for a thickness $d=3\lambda_2/4$ or $T=3\pi/2$, the loss could be determined easily. In handling dielectric materials with low loss one can obtain more accurate results than are obtainable from the chart, by using the following series approximation, valid in the range $C>1$ when $T\cong 3/2\pi$.

$$Te^{j\tau} = j \left(4.71 - \frac{0.212}{C} e^{j\tau} - \frac{0.0096}{C^2} e^{2j\tau} \right). \quad (22)$$

The complex dielectric constant can be found from γ_2 by the general relation

$$\gamma^2 = \left(\frac{2\pi}{\lambda_c} \right)^2 - \omega^2 \mu \epsilon^*. \quad (23)$$

The cut-off wave-length λ_c is determined by the

geometry of the pipe and the mode of propagation. The frequency of plane waves with a wave-length λ_c in the medium to which γ refers is called the cut-off frequency. The values of λ_c for the lowest order waves in different wave guides are

rectangular pipe ⁷	$\lambda_c = 2 \times \text{width}$
round pipe ⁸	$\lambda_c = 1.71 \times \text{diameter}$
coaxial line	$\lambda_c = \infty$

The value of γ_1 is given by

$$\gamma_1^2 = \left(\frac{2\pi}{\lambda_c} \right)^2 - \omega^2 \mu_0 \epsilon_0 = - \left(\frac{2\pi}{\lambda_1} \right)^2. \quad (24)$$

Bearing in mind that μ for dielectrics is generally μ_0 of free space, one obtains

$$\epsilon_2^* = \epsilon_0 \frac{(1/\lambda_c)^2 - (\gamma_2/2\pi)^2}{(1/\lambda_c)^2 + (1/\lambda_1)^2}. \quad (25)$$

For the coaxial line this equation reduces to

$$\epsilon_2^* = \epsilon_0 (\gamma_2/\gamma_1)^2. \quad (26)$$

The determination of ϵ_2^* may now be summarized as follows. The ratio $E_{\min}/E_{\max}$, the distances x_0 and d and the wave-length λ_1 are measured. From these measured values C and τ are calculated with the help of Eq. (20). From

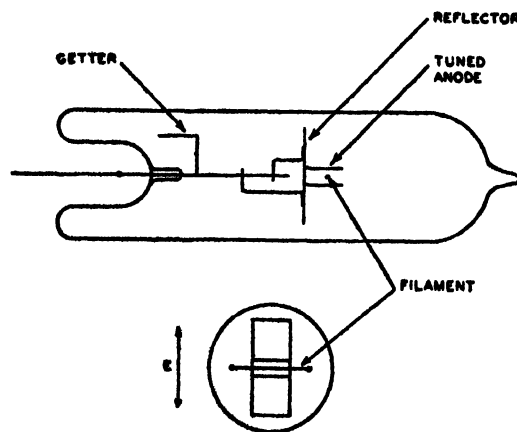


FIG. 6. Details of magnetron.

C and τ the values of T and τ are found from the chart or by a series approximation. γ_2 is found

⁷ L. J. Chu and W. L. Barrow, Proc. I.R.E. 26, 1520 (1938).

⁸ G. L. Southworth, Bell Sys. Tech. J. 15, 284 (1936); W. L. Barrow, Proc. I.R.E. 24, 1298 (1936).

by dividing $T\alpha''$ by d . Finally ϵ_1^* is determined from γ_2 by Eq. (25) or (26).

APPARATUS AND PROCEDURE

The apparatus used in 1940 is shown in Fig. 5. It is now obsolete in many details and will therefore be described only briefly.

A round wave guide P is mounted vertically and is fitted at the bottom with a detachable end piece in which either solid or liquid dielectrics may be inserted. A magnetron oscillator T with its magnet M is located at the upper end of the pipe. The output of the magnetron is monitored by a bolometer detector B . The standing wave above the dielectric is measured by means of the traveling crystal detector E which is driven by the lead screw W while its position is indicated to an accuracy of .001 inch by the revolution counter C .

The magnetron oscillator is of a special design shown in Fig. 6. An open anode construction, which serves to radiate the energy into the pipe, avoids the tuned parallel transmission line which, in conventional constructions,⁹ feeds the energy through the glass envelope to a dipole antenna. The principal advantage of this design is its simplicity of construction. It has an efficiency of about one percent and requires a critical adjustment of the magnetic field intensity and the orientation in the field.

The output of the magnetron is filtered and stabilized by means of a resonator R , a cylindrical tube fitting in the pipe and closed on both ends by diaphragms. The approximate dimensions of this filter for a wave-length of 6.0 cm are: length 3 inches, diameter of tube $1\frac{1}{2}$ inches and diameter

of hole in the diaphragm $\frac{1}{2}$ inch. The resonator R is inserted for part of its length in the pipe section holding the magnetron. The main pipe, which slides over the free end of the resonator, is adjusted to resonance and then clamped in position.

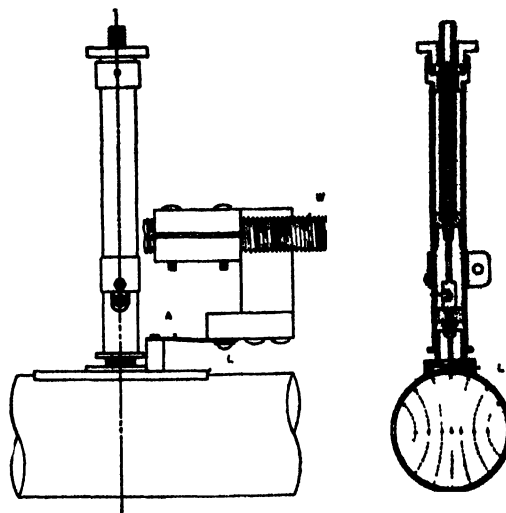


FIG. 8. Details of traveling crystal detector.

The construction and circuit diagram of the bolometer detector B are shown in Fig. 7. This arrangement proved very stable and troublefree, but of course less sensitive than a crystal detector of the silicon-tungsten type.

The slotted pipe maintains a wave which is polarized so that the electric field is most intense and the longitudinal magnetic field zero at the slot. The electric field is shown in Fig. 8 together with a more detailed drawing of the detector construction. Good metallic contact between detector housing and pipe is required and is provided by the slider L held firmly with the phosphor-bronze spring A .

In principle, one measurement with the dielectric material in the pipe would be sufficient. After tuning to resonance by adjusting the plunger D (Fig. 5), x_0 , $E_{\min}/E_{\max}$ and λ_1 are found by the rectified current and position of the traveling detector. In practice, a second measurement is carried through with the empty pipe for calibrating the detector. Here the field strength is known. It should go down to zero at the minimum and increase towards the maximum proportionally to the sine of the phase

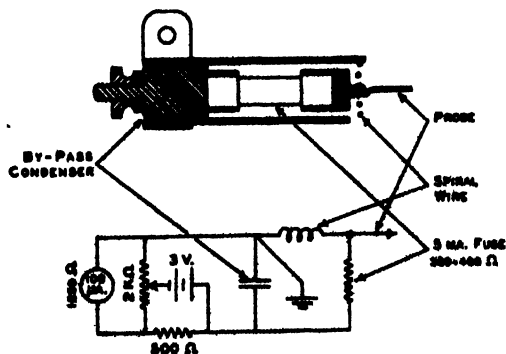


FIG. 7. Bolometer circuit and method of construction.

⁹ E. G. Linder, Proc. I.R.E. 27, 732 (1939).

angle θ , where

$$\theta = 2\pi(x - x_0)/\lambda_1. \quad (27)$$

Then

$$|E(x)_1| = E_{\max} |\sin \theta|. \quad (28)$$

The calibration curve of the detector is found by noting the detector current for several positions on both sides of the minimum. The amplitude at the minimum is an indication of attenuation in the pipe and frequency instability of the oscillator, which should be very low for proper operation. The wave-length is determined by measuring the positions of two minima.

With a low loss dielectric in the pipe, the measurements are carried out only in the neighborhood of the minimum, because the maximum is flatter and the detector is likely to be overloaded. In this case a simple formula can be derived for calculating the ratio of minimum to maximum field strength. The absolute value of electric field strength is found from (7).

$$\begin{aligned} |E(x)_1| &= |A_0| |1 + e^{-2(\gamma_1 x + \phi)}| \\ &= |A_0| |1 - e^{-2(\rho + j\theta)}|. \end{aligned} \quad (29)$$

Then the ratio of this field strength to the minimum, given by (11), is

$$\begin{aligned} \frac{|E(x)_1|}{E_{\min}} &= \frac{|e^{\rho+j\theta} - e^{-(\rho+j\theta)}|}{e^{\rho} - e^{-\rho}} = \frac{|\sinh(\rho + j\theta)|}{\sinh \rho} \\ &= \frac{|\sinh \rho \cos \theta + j \cosh \rho \sin \theta|}{\sinh \rho}. \end{aligned} \quad (30)$$

The square of this ratio is

$$\frac{|E(x)_1|^2}{E_{\min}^2} = \cos^2 \theta + \frac{\sin^2 \theta}{\tanh^2 \rho}. \quad (31)$$

Solving for $\tanh \rho$, one obtains the desired result.

$$\frac{E_{\min}}{E_{\max}} = \tanh \rho = \frac{\sin \theta}{\left(\frac{|E(x)_1|^2}{E_{\min}^2} - \cos^2 \theta \right)^{1/2}}. \quad (32)$$

The use of this indirect method of determining $E_{\min}/E_{\max}$ makes it possible to measure much

TABLE I. Dielectric constant and loss at $\lambda = 6.0$ cm. (Temperature $25 \pm 2^\circ\text{C}$ except where noted.)

Material	ϵ'/ϵ_0	$\tan \delta$
Solids		
Polystyrene	2.53	.0005 to .002
Lucite	2.57	.005 to .008
Plexiglas	2.60	.006
Cellulose acetate	3.24	.033
Polyethylene	2.25	.0002 to .0012
Glass		
Soda-lime-silicate	6.7	.014
Pyrex No. 774	4.8	.009
Pyrex No. 707	3.9	.0012
Fused quartz	3.80	.0001
Paraffin wax	2.25	.0002
Plywood (Bakelite bonded)	1.9	0.060
Poplar, grain $\parallel$ electric field	2.05	0.075
Poplar, grain $\perp$ electric field	1.75	0.065
Liquids		
Distilled water (28.5°C)	71	0.22
Glycerine U.S.P. (28°C)	5.6	0.60
Benzene (thiophene free)	2.25	<.001

smaller values of this ratio than could be done directly.

RESULTS

Table I gives the values of ϵ' and $\tan \delta$ at $\lambda = 6.0$ cm for a number of common materials including solids and liquids. The first measurements on most of these substances were made in 1940 on the apparatus described above. Since that time extensive and more accurate data at $\lambda = 3$ cm, $\lambda = 10$ cm and at longer wave-lengths have been obtained for a wide variety of dielectrics in this Laboratory and summarized in the two volumes *Tables of Dielectric Materials* prepared under O.S.R.D. Contract OEMsr-191. Table I has been edited to some extent on the basis of these more recent data.

ACKNOWLEDGMENT

The authors are indebted to the International Telephone and Telegraph Company for financial assistance in this research, and to Dr. H. H. Buttner, its Vice President, for his active interest in the progress of the work.

Initially, the study was to be in cooperation with Professor J. A. Stratton and Dr. L. J. Chu, of the Department of Physics, whose advice on theoretical questions was of great value.

Interpolation with the Aid of a Plot of First Differences

GORDON S. FULCHER

Washington, D. C.

(Received February 28, 1946)

Given the values of y_{-1} , y_0 , y_1 , and y_2 corresponding to x_{-1} , x_0 , x_1 , and x_2 , we first plot $\Delta y/\Delta x$, the ratio of first difference to abscissa interval, as a function of mean abscissa $\bar{x}$, and then draw a smooth curve through the graph points which gives the values of a function $S(\bar{x})$. To determine the value of y_p for any value of x_p between x_0 and x_1 , we substitute values of $S(\bar{x}')$ and $S(\bar{x}'')$, obtained from the curve, in the following equations: $y' = y_0 + f\Delta x \cdot S(\bar{x}')$; $y'' = y_1 - (1-f)\Delta x \cdot S(\bar{x}'')$; $y_p = y' + (y'' - y')(1+f)/3$; where $\bar{x}' = \frac{1}{2}(x_0 + x_p)$; $\bar{x}'' = \frac{1}{2}(x_1 + x_p)$ and $f = (x_p - x_0)/\Delta x$. With a carefully drawn smooth curve, the error in the result is only about one-sixth of the fourth difference term in the Newton-Beasel interpolation formula. In a simpler variation of the method, the function $L(\bar{x})$, whose values are given by straight lines connecting the graph points, is substituted for $S(\bar{x})$ in the equations; however, the accuracy of the result is only about one-sixth. It is shown

that the accuracy obtainable in any case can be estimated from the sagittas σ_1 and σ_0 for $\bar{x}$ equal to x_1 and x_0 , respectively, where $\sigma_1 = L(\bar{x}_1) - S(\bar{x}_1)$ and $\sigma_0 = L(\bar{x}_0) - S(\bar{x}_0)$, since the error is proportional to $(y'' - y')$ times $(\sigma_1 - \sigma_0)/(\sigma_1 + \sigma_0)$. This is true for unequal datum point intervals as well as for equal intervals. When the sagittas are equal or approximately so, the results may be as accurate if the rectilinear variation is used. Taking advantage of this fact, a diagram is given which enables the logarithm of any number to be determined to eight places by substituting values of $L(\bar{x})$, obtained by linear interpolation between given figures, in certain equations. The method has been found useful when a number of interpolations are required in a single set of data. An important advantage is that it can be used as easily when the abscissa intervals between datum points are unequal as when they are equal.

THE author became interested in methods of interpolation while making ballistic calculations during the winter of 1944-45 for the National Defense Research Committee. Accurate interpolation formulas have been available since the time of Newton; but not being a mathematician, I have a preference for graphical methods; so I amused myself investigating the use of a plot of first differences in connection with interpolation. This led to the method described below which has proved useful when a number of interpolations are required in a single set of data, such as can be represented by a single curve, and also when the interpolation is between datum points that are unevenly spaced.

TABLE I. Data used to illustrate the method.

Datum point	Abscissa x	Ordinate y	Difference Δy	$\Delta y/\Delta x$	Mean $\bar{x}$	Graph point
a	0	0				
b	5	3698	3698	739.6	2.5	P_1
c	10	6344	2646	529.2	7.5	P_2
d	15	8351	2007	401.4	12.5	P_3
e	20	9987	1636	327.2	17.5	P_4
f	25	11398	1411	282.2	22.5	P_5
g	30	12645	1247	249.4	27.5	P_6

1. DESCRIPTION OF THE METHOD

To explain the method we shall first apply it to an actual case with evenly spaced datum points, using the data given in Table I. To determine the values of y_p for a series of values of x_p between 10 and 20 we proceed as follows:

Step one: plotting first difference as a function of mean abscissa. If the abscissa interval Δx between datum points is constant, first difference Δy may be plotted as a function of mean abscissa $\bar{x}$; otherwise the ratio $\Delta y/\Delta x$ must be plotted. Figure 1 shows a plot of $\Delta y/\Delta x$ as a function of $\bar{x}$ for the graph points P_1 to P_6 whose co-ordinates are given in columns 5 and 6 of Table I. The original plot is 10 by 12 inches.

Step two: drawing smooth curve through the graph points. My procedure is first to draw a smooth curve through the points freehand, in pencil; next to draw through each point in ink a short segment of the curve with the help of French curves; and then to complete the curve between segments, again using French curves. Great care in drawing the curve is required only when maximum accuracy is desired. The smooth curve gives the values of a function of $\bar{x}$ which we shall call $S(\bar{x})$.

These two steps complete the graphical part

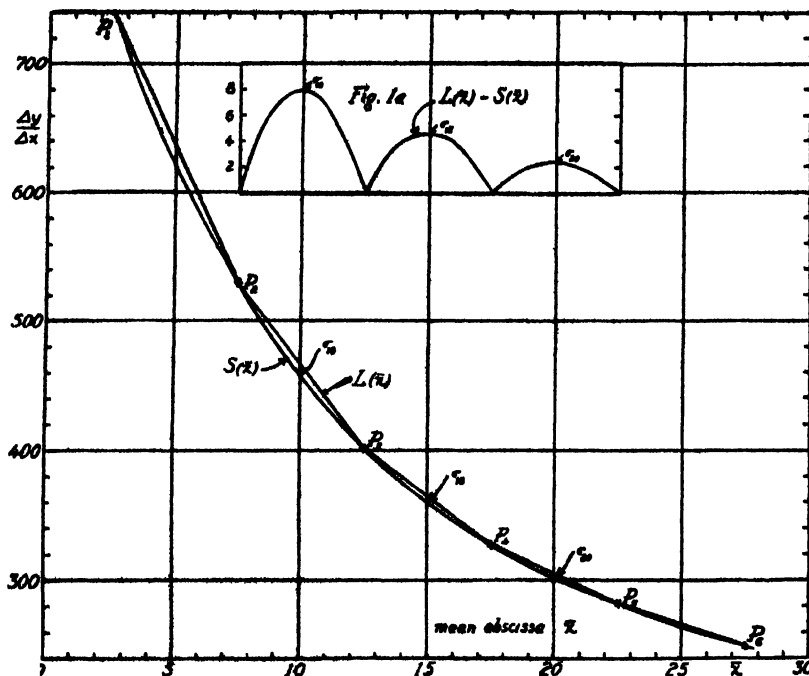


FIG. 1.

of the method. The straight lines shown connecting the graph points in Fig. 1 are used in connection with the rectilinear variation of the method which will be described later.

Step three: making the calculations. Assuming, for example, that we wish the value of y for x equal to 14, we make the following three simple calculations. First, using y_0 as starting point, we compute what we shall call y_{14s}' by adding $(14-10)$ times the value of $S(x)$ for x equal to $\frac{1}{2}(14+10)$, as read from the graph; then, using y_4 as starting point, we compute what we shall call y_{14s}'' by subtracting $(15-14)$ times the value of $S(x)$ for x equal to $\frac{1}{2}(15+14)$, as read from the graph; and finally we obtain a weighted average of the two approximate results by adding $(1+f)/3$ times their difference to y_{14s}' , where f is the ratio of $(14-10)$ to Δx . We thus obtain the following results:

$$y_{14s}' = 6344 + 4 \times 411.6 = 7990.4;$$

$$y_{14s}'' = 8351 - 1 \times 367.6 = 7983.4;$$

$$y_{14s} = 7990.4 + .6 \times (-7.0) = 7986.2.$$

As we shall see, this result happens to be correct within .3 unit. The subscript S is used to distinguish values obtained with the aid of the smooth curve from those obtained by use of the

straight lines connecting the graph points, as is done in the rectilinear variation of the method.

In general, the equations used in calculating the value of y_p corresponding to abscissa x_p between x_0 and x_1 are:

$$y_{ps}' = y_0 + (x_p - x_0) \cdot S(x') = y_0 + f \Delta x \cdot S(x'); \quad (1)$$

$$y_{ps}'' = y_1 - (x_1 - x_p) \cdot S(x'') \\ = y_1 - (1-f) \Delta x \cdot S(x''); \quad (2)$$

$$y_{ps} = y_{ps}' + (y_{ps}'' - y_{ps}') \cdot (1+f)/3; \quad (3)$$

where

$$x' = \frac{1}{2}(x_p + x_0); \quad (4)$$

$$x'' = \frac{1}{2}(x_1 + x_p); \quad (5)$$

and

$$f = (x_p - x_0)/(x_1 - x_0) = (x_p - x_0)/\Delta x. \quad (6)$$

To check the result obtained above for y_{14} we shall calculate the correct value by use of the Newton-Bessel formula¹ which may be written

¹The Newton-Bessel, Gregory-Newton, Lagrange, and other interpolation formulas are given in E. T. Whittaker and George Robinson's *Calculus of Observations* (Blackie and Son, London, 1940) and by other books on interpolation; tables giving the numerical values of the coefficients for the Gregory-Newton and Lagrange formulas are included in Herbert B. Dwight's *Mathematical Tables* (McGraw-Hill Book Company, Inc., New York, 1941) and in H. T. Davis' *Tables of Higher Mathematical Functions* (Principia Press, Inc., Bloomington, Indiana, 1935).

as follows:

$$y_p = \frac{1}{2}(y_0 + y_1) + (f - \frac{1}{2})\Delta y_1 - \frac{1}{2}f(1-f)(\Delta^2 y_0 + \Delta^2 y_1) - \frac{1}{3!}f(1-f)(f - \frac{1}{2})\Delta^3 y_1 \\ + \frac{1}{2 \times 4!}f(1-f)(f+1)(2-f)(\Delta^4 y_0 + \Delta^4 y_1) + \frac{1}{5!}f(1-f)(f+1)(2-f)(f - \frac{1}{2})\Delta^5 y_1 \\ - \frac{1}{2 \times 6!}f(1-f)(f+2)(f+1)(2-f)(3-f)(\Delta^6 y_0 + \Delta^6 y_1) - \text{etc.} \quad (7)$$

To apply this formula we need the differences given in Table II. Using these differences we obtain:

$$y_{14} = 7347.5 + 602.1 + 40.4 - 2.144 \\ - 1.922 + .020 - .048 = 7985.91.$$

The Lagrange seven-term formula gives 7985.92. Comparison of the value for y_{14} obtained above with these correct values shows an error of +.3 unit. Other values obtained with the aid of Fig. 1 are given in Table III together with accurate values given by the formula. The figures in the last column show that the maximum error comes out less than .5 unit in this case, that is less than $\Delta y/3000$.

To test the accuracy of the method further, four new plots of the data in Table I were made, and the smooth curves drawn through the graph points were used to repeat the calculations. For the 40 values thus obtained, including those in Table III, the average error came out less than .4 unit and in only two cases was the error more than 1 unit. The time taken to plot the six graph points and to draw the smooth curve was less than ten minutes in each case and only two French curves, not particularly suitable, were used.

An important advantage of the method is that it may be used when the datum points are not

TABLE II. Extension of Table I, giving second to sixth differences.

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$	$\Delta^6 y$
5	3698	-1052		-150		22	5
10	6344	- 639	413	-145	5	18	7.5
15	8351	- 371	268	-122	23	14	12.5
20	9987	- 225	146	- 85	37	10	17.5
25	11398	- 164	61	- 38	47	6	22.5
							25

evenly placed as easily as when Δx is constant. Before giving an example of such use, however, we shall describe the rectilinear variation of the method which is somewhat simpler, since it does not require the drawing of a smooth curve, and yet is sufficiently accurate for many cases.

TABLE III. Values obtained with the aid of Fig. 1.

x_p	y_{ps}'	y_{ps}''	$(y_{ps}' - y_{ps}'')$	Graphical result for y_p	Correct value of y_p	Error of y_{ps}
11	6789.2	6781.8	7.4	6786.2	6785.95	+.25
12	7210.8	7199.6	11.2	7205.6	7205.4	+.2
13	7610.9	7593.8	17.1	7605.0	7604.7	+.3
14	7990.4	7983.4	7.0	7986.2	7985.9	+.3
16	8703.8	8699.0	4.8	8701.9	8701.65	+.25
17	9042.6	9036.0	6.6	9039.5	9039.35	+.15
18	9368.6	9363.0	5.6	9365.6	9365.4	+.2
19	9683.8	9679.8	4.0	9681.4	9681.0	+.4

2. RECTILINEAR VARIATION OF THE METHOD

After plotting the graph points, we draw straight lines connecting them, as shown in Fig. 1. If we designate by $L(x)$ the function whose values are given by these lines, then the equations used in calculating y_{pL} are the same as Eqs. (1) to (6) except for the substitution of $L(x)$ for $S(x)$. That is, instead of Eqs. (1) to (3) we use the following:

$$y_{pL}' = y_0 + (x_p - x_0) \cdot L(x') = y_0 + f\Delta x \cdot L(x'); \quad (8)$$

$$y_{pL}'' = y_1 - (x_1 - x_p) \cdot L(x'') \\ = y_1 - (1-f)\Delta x \cdot L(x''); \quad (9)$$

$$y_{pL} = y_{pL}' + (y_{pL}'' - y_{pL}') \cdot (1+f)/3, \quad (10)$$

where x' , x'' , and f have the same meanings as above. For example, to determine y_{14} we make the following calculations:

$$y_{14L}' = 6344 + 4 \times 414.0 = 8000.0;$$

$$y_{14L}'' = 8351 - 1 \times 372.0 = 7979.0;$$

$$y_{14L} = 8000.0 + .6 \times (-210) = 7987.4.$$

The error of this result is +1.5 unit. Other results obtained with the aid of Fig. 1 are given in Table IV. The average error for these results is about six times the average error for the results given in Table III. As we shall see, this

TABLE IV. Results obtained using values of $L(x)$ from Fig. 1.

x_p	$(y_{pL}'' - y_{pL}')$	y_{pL}	Error of y_{pL}	Est. E_{pL}
11	-22.4	6787.6	+1.6	+2.3
12	-33.0	7208.2	+2.8	+3.3
13	-33.0	7607.4	+2.7	+3.3
14	-21.0	7987.4	+1.5	+2.3
16	-13.0	8702.8	+1.2	+1.6
17	-18.8	9041.5	+2.1	+2.3
18	-18.6	9367.7	+2.3	+2.3
19	-13.5	9682.9	+1.9	+1.6

is the ratio to be expected if the smooth curve is accurately drawn.

This rectilinear variation does not require a plot since the values of $L(x')$ and $L(x'')$ can be calculated by use of the following equations:

$$\Delta x \cdot L(x') = \Delta y_1 - \frac{1}{2}(1-f) \cdot \Delta^2 y_0; \quad (11)$$

$$\Delta x \cdot L(x'') = \Delta y_1 + \frac{1}{2}f \cdot \Delta^2 y_1. \quad (12)$$

However, when a number of interpolations are to be made, a plot is useful.

3. GRAPHICAL METHOD OF ESTIMATING THE ACCURACY OF RESULTS

If the smooth curve is drawn and also the chords connecting the graph points, as in Fig. 1, the accuracy obtainable by this method of interpolation can be estimated as follows. The error of y_{pL} is approximately given by the following equation:

$$E_{pL} = -\frac{3(\sigma_1 - \sigma_0)}{8(\sigma_1 + \sigma_0)}(y_{pL}'' - y_{pL}'), \quad (13)$$

where σ_1 represents the vertical sagitta at x_1 , that is the vertical distance between the chord P_2P_4 and the smooth curve at x equal to x_1 , and σ_0 represents the corresponding vertical sagitta at x_0 . Evidently,

$$\sigma_1 = L(x_1) - S(x_1), \quad (14)$$

and

$$\sigma_0 = L(x_0) - S(x_0). \quad (15)$$

According to Fig. 1a, σ_{30} , σ_{15} , and σ_{10} are equal to

about 2.3, 4.6, and 8 units, respectively. Using these values, we obtain for E_{pL} the estimates given in the last column of Table IV. These estimates are seen to agree approximately with the actual errors given in the preceding column.

If the smooth curve is accurate, the error of y_{pS} will be about one-sixth of E_{pL} . With an inaccurate curve, the ratio may be greater or less than six. In the case of Fig. 1, in spite of the fact that the values of $S(x)$ given by the curve are from .3 to .8 unit too large, the errors in the values of y_{pS} happen to come out somewhat less than if the smooth curve had been accurate, averaging +.27 instead of +.37 unit. (See Table X, below.) Thus, even with a hastily drawn smooth curve, the values of y_{pS} obtained with the aid of it are likely to be several times as accurate as values of y_{pL} obtained from the chords.

An interesting case is when the two sagittas are approximately equal. In this case, according to Eq. (13), the error of results obtained from values of $L(x)$ will be very small, and there is no point in using the smooth curve except to ascertain that the sagittas are equal. Therefore when the sagittas are found to be nearly equal accurate values of $L(x)$, calculated by use of Eqs. (11) and (12), should be used to determine y_{pL} if maximum accuracy is desired.

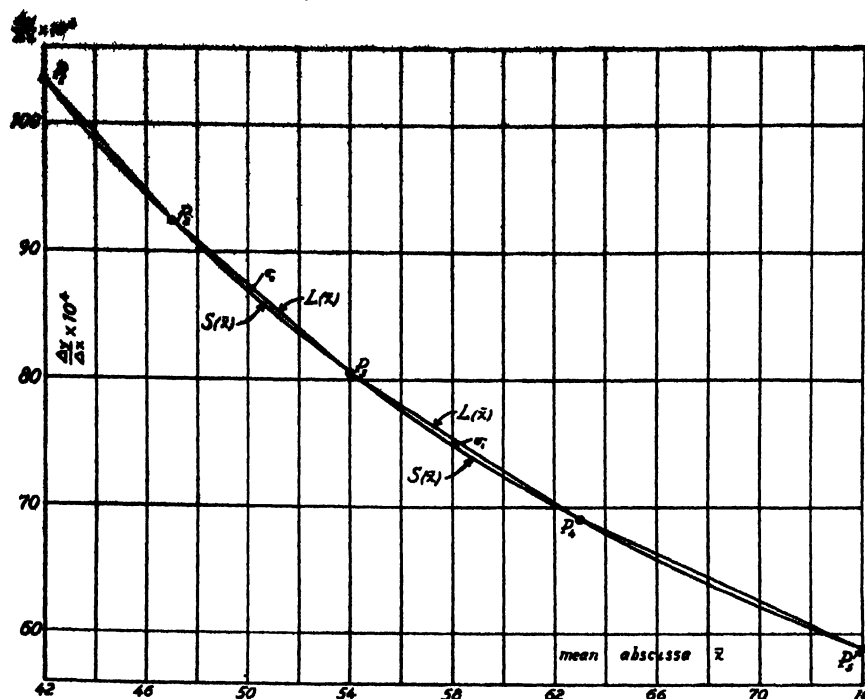
4. CASE OF UNEQUAL INTERVALS BETWEEN DATUM POINTS

In the example used above to illustrate the method, the abscissa interval Δx between datum points is constant. We shall now apply the method to a case where the abscissa intervals are unequal, using the data given in Table V.

TABLE V. Data used to illustrate the method as applied to unequally spaced datum points. ($y = \log x$.)

x	y	Δy	Δx	$\Delta y/\Delta x$	Mean x	Graph point
40	1.60206,0					
		.04139,3	4	.01034,82	42	P_1
44	1.64345,3					
		.05551,7	6	.00925,29	47	P_2
50	1.69897,0					
		.06445,8	8	.00805,72	54	P_3
58	1.76342,8					
		.06908,1	10	.00690,81	63	P_4
68	1.83250,9					
		.07056,1	12	.00588,17	74	P_5
80	1.90309,0					

FIG. 2.



Whereas when Δx is constant either Δy or $\Delta y/\Delta x$ may be plotted as a function of mean x , when Δx is not constant it is necessary, of course, to plot $\Delta y/\Delta x$. Figure 2 shows such a plot of the data in Table V. The original is about 10 by 13 inches. Using the smooth curve drawn through the graph points, values of y_{ps} were determined for integral values of x_p between 50 and 58. The resulting values of $(y_{ps}'' - y_{ps}')$ and of E_s are given in Table VI. The maximum error came out only 0.5×10^{-3} less than $\Delta y/10,000$. Another similar plot, made independently, gave the same maximum error.

Much to my surprise, the results obtained by use of the chords between the datum points came out as accurate as those obtained with the

smooth curve. Then, in place of the values of $L(\bar{x})$ read from the chords, I tried values accurately calculated by use of Eqs. (11) and (12). The resulting values of $(y_{pL}'' - y_{pL}')$ and of E_L are given in Table VI. The maximum error came out only -0.21×10^{-3} , less than $\Delta y/30,000$. Then I calculated E_L by use of Eq. (13), substituting for σ_1 and σ_0 the values 4.3 and 4.1, respectively, as determined from Fig. 2. For x equal to 54 the estimated value for E_L comes out $(y_{pL}'' - y_{pL}')/112$ or -0.31×10^{-3} . This is as close to the actual value as is to be expected.

As will be described later, investigation of other series of uneven intervals shows that with uneven datum point intervals as with constant intervals, Eq. (13) gives, at least approximately, the value of E_L to be expected. From this it follows that if when interpolating in a certain interval, the adjacent intervals can be adjusted so that the two sagittas come out equal, then the interpolation results obtained by using accurate values of $L(\bar{x})$ will be unusually accurate. This principle has been used to construct a diagram, see Fig. 5 below, which enables the logarithm of any number to be determined to six to eight places, from the data on the diagram, by interpolation, using the rectilinear method.

TABLE VI. Results for log 51 to log 57 from data in Table V. (Unit is .00001.)

x	$(y_{ps}'' - y_{ps}')_{\text{(Fig. 2)}}$	$E_s_{\text{(Fig. 2)}}$	$(y_{pL}'' - y_{pL}')_{\text{(accurate)}}$	$E_L_{\text{(accurate)}}$
51	-4.1	.0	-15.09	-.21
52	-6.2	+1	-25.87	-.18
53	-7.6	+3	-32.34	-.13
54	-8.2	-.5	-34.49	-.08
55	-8.2	-.3	-32.34	-.06
56	-5.8	-.4	-25.87	-.06
57	-3.1	.0	-15.09	-.10
Average:	-6.2	$\pm .2$	-25.87	-.12

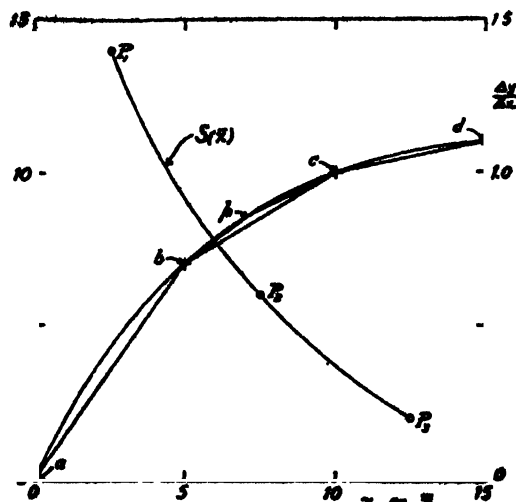


FIG. 3. Plot of a function of x and of $S(x)$, a function of $\bar{x}$.

The same method can be applied to any function for which the values of $\Delta y/\Delta x$ do not vary too widely, for example to sine θ or cosine θ or to tangent of 0° to 45° .

5. INVESTIGATION WHICH LED TO THE INTERPOLATION METHOD

We shall assume that we are given the coordinates of the datum points a , b , c , and d of the function of x plotted in Fig. 3 and that we wish to determine the value of y_p at a point p with abscissa x_p . We plot the datum points, draw a smooth curve through them in order to locate p approximately, and draw the chords between datum points and finally the chords bp and pc . Now since

$$y_p = y_b + (x_p - x_b) \cdot (\text{slope of chord } bp),$$

and

$$y_p = y_c - (x_c - x_p) \cdot (\text{slope of chord } pc), \quad (16)$$

if we knew the slope of bp or of pc we could determine y_p . It occurred to me to investigate whether these slopes can be determined from the known slopes of the chords between datum points.

If the point p is moved along the curve from a to b , the slope of pb varies continuously from that of ab to that of the tangent to the curve at b ; if p is then moved on to c , the slope of bp varies continuously from that of the tangent at b to that of the chord bc , while the slope of pc varies continuously from that of bc to that of

the tangent at c ; and if p is then moved to d , the slope of cp varies from that of the tangent at c to that of cd . Meanwhile, the mean abscissa of the chord to or from p increases from $\frac{1}{2}(x_a + x_b)$ to $\frac{1}{2}(x_c + x_d)$. From this it is evident that the slope of the chord between p and one or other of the two nearest datum points, considered always as extending to the right, is a continuous function of the mean abscissa $\bar{x}$. If we designate this function by $F(\bar{x})$, then Eqs. (16) become

$$y_p = y_b + (x_p - x_b) \cdot F(\bar{x}')$$

$$= y_c - (x_c - x_p) \cdot F(\bar{x}''), \quad (17)$$

where $\bar{x}'$ is the mean of x_p and x_b , and $\bar{x}''$ is the mean of x_p and x_c .

In Fig. 3 the slopes of the chords ab , bc , and cd , that is the values of $\Delta y/\Delta x$ for the three intervals, are plotted as a function of mean abscissa and a smooth curve is drawn through the graph points. If we designate by $S(\bar{x})$ the function of $\bar{x}$ represented by this curve, then $S(\bar{x})$ has the same values as $F(\bar{x})$ for $\bar{x}$ equal to the mean abscissas of the chords ab , bc , and cd .

To investigate the relation between $S(\bar{x})$, whose values can be determined graphically with more or less accuracy from the data, and $F(\bar{x})$, whose values we wish for substitution in Eq. (17), the function $y = \log x$ was chosen for study since it is an unusually difficult one for accurate interpolation. Given the data included in Table VII, the first problem was to determine the relation between $S(\bar{x})$ and $F(\bar{x})$ for this function.

Using a seven-place table of logarithms, the values of $F(\bar{x})$ were computed for all integral and half-integral values of $\bar{x}$ between 45 and 70. The values so obtained were plotted in Fig. 4

TABLE VII. Data for the function $y = \log x$.

Datum point	x	y	$\Delta y/\Delta x$	$\bar{x}$	Graph point
a	30	1.47712,12			
b	40	1.60206,00	.01249,39	35	P_{35}
c	50	1.69897,00	.00969,10	45	P_{45}
d	60	1.77815,12	.00791,81	55	P_{55}
e	70	1.84509,80	.00669,47	65	P_{65}
f	80	1.90309,00	.00579,92	75	P_{75}

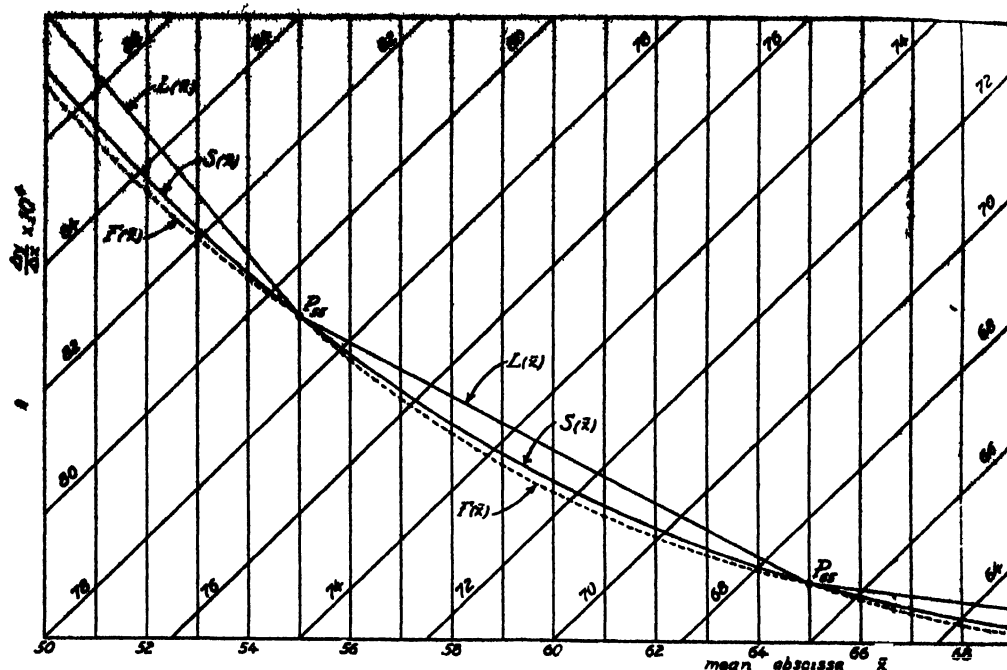


FIG. 4.

and a dashed curve drawn through them. The original plot is 10 by 15 inches; and in order to increase the vertical scale, lines of constant ordinate were drawn in at 45° to the lines of constant abscissa. The curve for $F(x)$ passes through the graph points, of course.

A smooth curve through the graph points will obviously lie near the $F(x)$ curve. To investigate the relation between these functions, the smooth curve shown in Fig. 4 was drawn through points whose ordinates were calculated by use of the following equation:

$$10 \cdot S(x) = \log(x+5) - \log(x-5).$$

This was done to eliminate effects of inaccuracy in drawing the curve. Finally, the graph points were connected by straight lines which give the values of the function $L(x)$. It then became evident that the $S(x)$ curve lies everywhere between the functions $F(x)$ and $L(x)$ and the difference $S(x)$ minus $F(x)$ is about one-third of the difference $L(x)$ minus $S(x)$ in this case. It was found that for x between 50 and 60 the ratio of the differences is in fact close to 3 on the average, being equal to 3.12. This suggested that if we subtract from $S(x)$ one-third of $[L(x) - S(x)]$ we shall obtain a function of x whose values are approximately equal to $F(x)$.

If we call this function $C(x)$, then

$$C(x) = S(x) - \frac{1}{3}[L(x) - S(x)]; \quad (18)$$

and if we substitute $C(x)$ for $F(x)$ in Eq. (17), we shall obtain approximately correct values for y_p .

Since the error in the result obtained is $(x_p - x_b)$ or $(x_o - x_p)$ times the error in $C(x')$ or in $C(x'')$, we shall reduce the error on the average if we write two equations, as follows,

$$y_{pc'} = y_b + (x_p - x_b) \cdot C(x') = y_b + f \Delta x \cdot C(x'); \quad (19)$$

$$y_{pt''} = y_o - (x_o - x_p) \cdot C(x'')$$

$$= y_o - (1 - f) \Delta x \cdot C(x''); \quad (20)$$

and use Eq. (19) for cases where f is between 0 and $\frac{1}{2}$, and Eq. (20) for cases where f is greater than $\frac{1}{2}$.

Results for $\log 51$ to $\log 59$, obtained by substituting values of $C(x)$ from Fig. 4 in Eqs. (19) and (20), are given in Table VIII. The average error comes out about $.5 \times 10^{-4}$, or $\Delta y / 16,000$. This shows that the method is sufficiently accurate providing an accurate smooth curve can be drawn through the graph points.

To determine whether the relation between the functions $F(x)$, $L(x)$, and $S(x)$ suggested by

TABLE VIII. Results for log 51 to log 59, obtained by substituting values of $C(x)$ from Fig. 4 in Eqs. (19) and (20). (Unit is .00001.)

x_p	x	$F(x)$	$C(x)$	E_C'	E_C''
51	50.5	860.0	859.9	-.1	
52	51	851.7	851.5	-.3	
53	51.5	843.5	843.3	-.7	
54	52	835.6	835.4	-.8	
55	52.5	827.8	827.6	-1.3	
55	57.5	755.8	755.9		-.7
56	58	749.1	749.2		-.5
57	58.5	742.5	742.6		-.2
58	59	736.2	736.2		-1.1
59	59.5	729.9	729.9		.0

Fig. 4 is a general relation, equations for these functions in terms of the differences were derived from the Newton-Bessel formula, Eq. (7). From these equations, which are given in Appendix A, it follows that

$$\Delta x \cdot [C(x') - F(x')] = -\frac{f(1-f)(1+f)}{144} (\Delta^4 y_0 + \Delta^4 y_1) + \text{etc.} \quad (21)$$

$$\Delta x \cdot [C(x'') - F(x'')] = +\frac{f(1-f)(2-f)}{144} (\Delta^4 y_0 + \Delta^4 y_1) - \text{etc.} \quad (22)$$

In absolute value these differences each amount to less than one fourth of the fourth difference term in the Newton-Bessel formula, Eq. 7. In our example above, since $(\Delta^4 y_0 + \Delta^4 y_1)$ equals $-.00702$, the value of these differences for $f = \frac{1}{2}$ should be about $\pm 1.8 \times 10^{-5}$ divided by Δx , that is $\pm .18 \times 10^{-4}$. The differences between columns 3 and 4 of Table VIII agree with this result.

When an inaccurate smooth curve is used in obtaining the values of $C(x)$ for substitution in Eq. (19) or (20), however, any error in $S(x)$ is multiplied by $(4/3)(x_p - x_0)$ or $(4/3)(x_s - x_p)$ so that in the case of Fig. 1, for example, the results for y_{ps} came out with errors which average four times the average error for y_{pc} when accurate values of $S(x)$ were used. (See Table X, below.)

It occurred to me that if values of $S(x)$ are substituted in Eqs. (19) and (20) for corresponding values of $C(x)$, since $S(x')$ is greater than $C(x')$ if $S(x'')$ is greater than $C(x'')$, the errors in the two results obtained, y_{ps}' and y_{ps}'' , will be opposite in sign; therefore, a weighted mean

of the two values might be fairly accurate. To test this, accurate values of $S(x)$ from Fig. 4 were substituted in Eqs. (1) and (2), and values for y_{ps}' and y_{ps}'' were obtained for integral values of x_p between 50 and 60. From these, the figures in column 2 of Table IX were obtained. Figures for the difference between y_{ps}' and the correct values are given in column 3, and the ratios of these differences to the differences in the second column are given in column 4. A plot of this ratio as a function of f showed that the values run parallel to the values of $(1+f)/3$ and differ from them by only about .067. It follows that, approximately,

$$y_p = y_{ps}' + (y_{ps}'' - y_{ps}')(1+f)/3. \quad (23)$$

If we call this result y_{ps} , then Eq. (23) is the same as Eq. (3); hence we have arrived at the method described at the beginning of the article.

Equation (23) is equivalent to the following equation:

$$y_{ps} = \left[\frac{1}{2} - \frac{(f-\frac{1}{2})}{3} \right] \cdot y_{ps}' + \left[\frac{1}{2} + \frac{(f-\frac{1}{2})}{3} \right] \cdot y_{ps}''. \quad (24)$$

The value y_{ps} is thus a weighted average of y_{ps}' and y_{ps}'' . It was found that the error of this average is less, with an inaccurate curve for $S(x)$, than the error of the corresponding result y_{pc} obtained from the function $C(x)$, on the average. This conclusion is illustrated by the figures given in Table X for the errors in the results obtained, by use of various equations, from the data in Table I and from Fig. 1. Although with an accurate smooth curve E_C averages only half E_S , with the actual curve for $S(x)$, which gives values which are up to .8 unit too large, E_C averages three times E_S . Of course,

TABLE IX. Results calculated from $S(x)$ in Fig. 4. (Unit is .00001.)

x_p (1)	$(y_{ps}'' - y_{ps}')^2$ (2)	$(y_p - y_{ps}')^2$ (3)	Ratio of (3) to (2)	$\frac{1}{2}(1+f)$
51	- 6.45	-2.80	.434	.367
52	-11.25	-5.27	.468	.400
53	-14.49	-7.28	.502	.433
54	-16.26	-8.70	.535	.467
55	-16.62	-9.44	.568	.500
56	-15.67	-9.40	.600	.533
57	-13.47	-8.51	.632	.567
58	-10.08	-6.68	.663	.600
59	- 5.57	-3.86	.693	.633

this is only one case. However, my experience shows that Eqs. (1) to (3) will give more accurate values than Eqs. (19) and (20), as a rule. Nevertheless, since only one of the two Eqs. (19) and (20) need be used in any particular case, it may be convenient to use the $C(x)$ function rather than the $S(x)$ function when this gives sufficient accuracy.

Table X also gives the values of E_L , the errors of results obtained by substituting values of $L(x)$ in Eqs. (8) to (10). When accurate values of $L(x)$ are used,

$$E_L = -\frac{f(1-f)(f+1)(2-f)}{48} \times (\Delta^4 y_0 + \Delta^4 y_1) - \text{etc.}; \quad (25)$$

and since

$$(y_{pL}'' - y_{pL}') = -\frac{f(1-f)}{2} \Delta^3 y_1; \quad (26)$$

and

$$\frac{(\sigma_1 - \sigma_0)}{(\sigma_1 + \sigma_0)} = -\frac{1}{4} \frac{(\Delta^4 y_0 + \Delta^4 y_1)}{\Delta^3 y_1} - \text{etc.}, \quad (27)$$

where σ_1 and σ_0 are the vertical sagittas defined above by Eqs. (14) and (15), therefore, approximately,

$$E_L = -\frac{3(\sigma_1 - \sigma_0)}{8(\sigma_1 + \sigma_0)} (y_{pL}'' - y_{pL}'). \quad (28)$$

This is the same as Eq. (13) above. In the case of Fig. 4, σ_1 is equal to 51.33 and σ_0 to 89.55; hence E_L should be equal to about $.1017 \times (y_{pL}'' - y_{pL}')$ for values of x_p between 50 and 6.0

TABLE X. Errors of results from data in Table I, (1) using values of the functions from Fig. 1, and (2) using accurate values.

x_p	(1) Values from Fig. 1			(2) Accurate values		
	E_L	E_G	E_S	E_L	E_G	E_S
11	+1.6	+.8	+.3	+2.09	-.15	+.34
12	+2.8	+1.2	+.2	+3.07	-.3	+.30
13	+2.7	-2.2	+.3	+3.03	-.25	+.47
14	+1.5	-1.0	+.3	+1.96	-.1	+.33
16	+1.2	+.7	+.25	+1.55	-.1	+.27
17	+2.1	+.8	+.15	+2.40	-.25	+.42
18	+2.3	+.4	+.2	+2.30	-.15	+.40
19	+1.9	-.5	+.4	+1.47	-.05	+.22
Average:	+2.0	±.9	+.3	+2.22	-.17	+.37

TABLE XI. Results for log 54 interpolated in several series of datum points with various intervals.

Absolute of datum points	Series of intervals	$E_{4L} \times 10^3$	$\frac{E_{4L}}{(y_{4L}'' - y_{4L}')} \times 10^3$	$\frac{3(\sigma_1 - \sigma_0)}{3(\sigma_1 + \sigma_0)}$	f
42, 44, 46, 48, 50, 52, 54	2-2-2-12-12	-2.08	+.061	+.080	-.3
42, 44, 50, 52, 54, 56	4-2-2-10-12	-0.08	+.008	+.015	+.1
57, 42, 50, 52, 54, 57, 77	6-7-2-0-10	+1.44	-.041	-.080	+.3
54, 44, 50, 52, 54, 74	8-2-2-2-2	+3.10	-.088	-.088	+.5
52, 40, 50, 52, 54, 56	12-10-2-0-4	+5.40	-.165	-.165	+.8

Actually the average ratio of E_L to $(y_{pL}'' - y_{pL}')$ comes out .101.

With an accurate smooth curve, E_S is about one sixth of E_L . According to the figures in Table X, the ratio is one to six for accurate values and about one to six for values from Fig. 1.

In the case of unequal intervals between datum points, after the results given in Table VI had been obtained, various other series of unequal intervals were investigated. Accurate values of $L(x)$ were used. The results for E_L for f equal to $\frac{1}{2}$, that is for log 54, are given in Table XI; also values of the ratio of E_{4L} to $(y_{4L}'' - y_{4L}')$ and of $\frac{3(\sigma_1 - \sigma_0)}{3(\sigma_1 + \sigma_0)}$. The agreement between the figures in columns 4 and 5 of Table XI indicates that Eq. (28) gives the value of E_L whether the intervals between datum points are equal or unequal as accurately as the ratio of $(\sigma_1 - \sigma_0)$ to $(\sigma_1 + \sigma_0)$ can be determined graphically.

The values of E_S given in the last column of Table XI were calculated by use of as accurate a smooth curve as could be obtained but are only approximately correct. They indicate that the ratio of E_L to E_S is practically equal to six except when the errors approach zero. The general conclusion is that the greater the value of the ratio of $(\sigma_1 - \sigma_0)$ to $(\sigma_1 + \sigma_0)$ in absolute value, the greater the advantage in accuracy of using the smooth curve rather than the chords between graph points in interpolation by this method.

6. DIAGRAM WHICH ENABLES ACCURATE VALUES TO BE OBTAINED BY LINEAR INTERPOLATION

For a certain interval, say log 50 to log 60, accurate values can be obtained by linear interpolation between graph points, as we have seen, if adjacent datum intervals are correctly chosen. Now the effect of changing these intervals is to

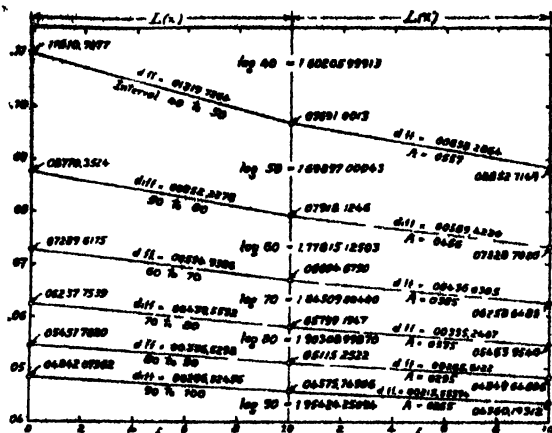


FIG. 5. Values of $10L(x')$ and of $10L(x'')$ and other data for use in calculating $\log x$.

change the slopes of the lines connecting the graph point at the center of the interval with the adjacent graph points. These slopes are equal to $[L(x)_1 - L(x)]/\frac{1}{2}\Delta x$ and $[L(x)_0 - L(x)]/\frac{1}{2}\Delta x$, where $L(x)_1$, $L(x)$, and $L(x)_0$ represent the values of $L(x)$ for x equal to x_1 , x , and x_0 , respectively. Given the value of $L(x)$, we can adjust the values of $L(x)_1$ and of $L(x)_0$ so that (1) the error of y_{1L} is zero, and (2) so that the error for f equal to some value, say .2, is the same as the error for f equal to say .8. The equations for $L(x)_1$ and $L(x)_0$ are then as follows:

$$\Delta x[L(x)_1 + L(x)_0] = 2\Delta y_1 - \frac{(y_8 - y_2) - (.8 - .2) \cdot \Delta y_1}{2 \times .8 \times .2 \times (.8 - .5)/3}; \quad (29)$$

$$\Delta x[L(x)_1 - L(x)_0] = 4(y_1 + y_0 - 2y_4); \quad (30)$$

from which $2L(x)_1$ can be obtained by adding Eqs. (29) and (30), and $2L(x)_0$ by subtracting Eq. (30) from (29). Substituting $\log 50$, $\log 52$, $\log 55$, $\log 58$, and $\log 60$ for y_0 , y_2 , y_4 , y_8 , and y_1 , respectively, we obtain:

$$\begin{aligned} 10[L(x)_1 + L(x)_0] &= .15836,25006 - (.04742,46500 \\ &\quad - .04750,87476)/.032; \\ &= .15836,25006 + .00262,8050 \\ &= .16099,05506. \\ 10[L(x)_1 - L(x)_0] &= 4(3.47712,12546 \\ &\quad - 3.48072,53790) \\ &= -.01441,64976. \end{aligned}$$

Hence:

$$L(x)_1 = .00732,87026; \text{ and } L(x)_0 = .00877,03524.$$

If we substitute these values in the equations:

$$L(x') = L(x)_1 + (1-f)[L(x)_0 - L(x)_1]; \quad (31)$$

$$L(x'') = L(x)_1 - f[L(x)_1 - L(x)_0]; \quad (32)$$

and substitute the values of $L(x')$ and $L(x'')$ so obtained in Eqs. (8), (9), and (10), we shall obtain values of $\log x_p$ for values of x_p between 50 and 60, with a maximum error of .00000,18; and if we substitute for Eq. (10) the following:

$$y_p = y_{pL}' + (y_{pL}'' - y_{pL}') \times \left\{ \frac{1}{2} + \frac{(f - \frac{1}{2})}{3} - A \cdot (f - \frac{1}{2})^2 \right\}, \quad (33)$$

where in this case A has the value 0.0456, the maximum error will be reduced to about 0.00000,005.

The constants in Table XII enable the logarithm of any number to be calculated. The result will be accurate to from five to seven places, depending on which interval is used, if Eq. (10) is used, and to two places more if Eq. (33) is used. For example, in computing $\log 48$ we get $y_{48}' = 1.68169,9564$; $y_{48}'' = 1.68092,9260$; $(y_{48}'' - y_{48}') = -.00078,0304$; $y_{48} = 1.68123,7382$, if Eq. (10) is used, and $y_{48} = 1.68124,1243$ if Eq. (33) is used. The errors of the two results are $-.3856 \times 10^{-8}$ and $+.0006 \times 10^{-8}$, respectively. If $\log 48$ is obtained by computing $\log 96$ and subtracting $\log 2$ the corresponding errors come out -3.36×10^{-8} and -0.16×10^{-8} , respectively. Thus, by using the interval 80 to 90 or 90 to 100, logarithms accurate to eight places of decimals can be obtained from the data in Table XII.

TABLE XII. Constants for calculating $\log x$ by use of function $L(x)$.^a

x_0 and x_1	$10L(x)_1$	$10L(x)_0$	$10L(x)$
40	1.60205,9991	.09991,0013	.11010,7277
50	1.69897,0004	.07918,1246	.08770,3524
60	1.77815,1250	.04694,6790	.07289,6175
70	1.84509,8040	.05799,1947	.06237,7539
80	1.90308,9987	.06115,2323	.05451,7820
90	1.95424,2509	.04842,0736	.04549,6400
100	2.00000,0000	.04360,1931	.04360,1931

^a The equations to be used in calculating $\log x$ are: Eqs. (31), (32), (8), (9), and either (10) or (33).

Although a graph is not necessary, such a diagram as that in Fig. 5 is useful as a guide to the algebraic signs to use in computing $L(\bar{x})$ and $L(\bar{x}'')$; moreover, it gives all the data included in Table XII.

APPENDIX. EQUATIONS FOR FUNCTIONS OF MEAN $\bar{x}$

The datum points are equally spaced in all cases; and the functions are all accurate. The equations are derived from the Newton-Bessel formula, Eq. (7) above.

$$\Delta x \cdot L(\bar{x}) = \Delta y_1 - \frac{(1-f)}{4} (\Delta^2 y_0 + \Delta^2 y_1) + \frac{(1-f)}{4} \Delta^3 y_1; \quad (34)$$

$$\Delta x \cdot L(\bar{x}'') = \Delta y_1 + \frac{f}{4} (\Delta^2 y_0 + \Delta^2 y_1) + \frac{f}{4} \Delta^3 y_1; \quad (35)$$

$$\begin{aligned} \Delta x \cdot S(\bar{x}) = \Delta y_1 - \frac{(1-f)}{4} (\Delta^2 y_0 + \Delta^2 y_1) + \frac{(1-f)^2}{8} \Delta^3 y_1 \\ + \frac{(1-f)(1+f)(3-f)}{4^2 \times 3!} (\Delta^4 y_0 + \Delta^4 y_1) - \frac{(1-f)^2(1+f)(3-f)}{4^2 \times 4!} \Delta^5 y_1 - \text{etc.} \end{aligned} \quad (36)$$

$$\Delta x \cdot S(\bar{x}'') = \Delta y_1 + \frac{f}{4} (\Delta^2 y_0 + \Delta^2 y_1) + \frac{f^2}{8} \Delta^3 y_1 - \frac{f(2+f)(2-f)}{4^2 \times 3!} (\Delta^4 y_0 + \Delta^4 y_1) - \frac{f^2(2+f)(2-f)}{4^2 \times 4!} \Delta^5 y_1 + \text{etc.} \quad (37)$$

$$\begin{aligned} \Delta x \cdot F(\bar{x}) = \Delta y_1 - \frac{(1-f)}{4} (\Delta^2 y_0 + \Delta^2 y_1) - \frac{(1-f)(f-\frac{1}{2})}{3!} \Delta^3 y_1 \\ + \frac{(1-f)(1+f)(2-f)}{2 \times 4!} (\Delta^4 y_0 + \Delta^4 y_1) + \frac{(1-f^2)(2-f)(f-\frac{1}{2})}{5!} \Delta^5 y_1 - \text{etc.} \end{aligned} \quad (38)$$

$$\begin{aligned} \Delta x \cdot F(\bar{x}'') = \Delta y_1 + \frac{f}{4} (\Delta^2 y_0 + \Delta^2 y_1) + \frac{f(f-\frac{1}{2})}{3!} \Delta^3 y_1 \\ - \frac{f(f+1)(2-f)}{2 \times 4!} (\Delta^4 y_0 + \Delta^4 y_1) - \frac{f(1+f)(2-f)(f-\frac{1}{2})}{5!} \Delta^5 y_1 + \text{etc.} \end{aligned} \quad (39)$$

$$E_L = -\frac{f(1-f)(1+f)(2-f)}{48} \left[(\Delta^4 y_0 + \Delta^4 y_1) + \frac{2}{5} (f-\frac{1}{2}) \Delta^5 y_1 \right] - \text{etc.} \quad (40)$$

$$E_S = -\frac{f(1-f)(1+f)(2-f)}{288} \left[(\Delta^4 y_0 + \Delta^4 y_1) + \frac{9}{10} (f-\frac{1}{2}) \Delta^5 y_1 \right] + \text{etc.} \quad (41)$$

$$E_{C'} = -\frac{f^2(1-f^2)}{144} (\Delta^4 y_0 + \Delta^4 y_1) + \text{etc.} \quad (42)$$

$$E_{C''} = +\frac{f(1-f)^2(2-f)}{144} (\Delta^4 y_0 + \Delta^4 y_1) + \text{etc.} \quad (43)$$

$$(y_L'' - y_L') = -\frac{f(1-f)}{2} \Delta^3 y_1; \quad (44)$$

$$(y_{s''} - y_{s'}) = -\frac{f(1-f)}{8} \Delta^2 y_1 - \frac{f(1-f)(f-\frac{1}{2})}{48} (\Delta^4 y_0 + \Delta^4 y_1) + \text{etc.} \quad (45)$$

$$\sigma_1 = -\frac{1}{8} \Delta^2 y_1 + \frac{1}{16} \Delta^4 y_1 - \frac{3}{128} \Delta^6 y_1 + \frac{3}{256} \Delta^8 y_1 - \text{etc.} \quad (46)$$

$$\sigma_0 = -\frac{1}{8} \Delta^2 y_1 - \frac{1}{16} \Delta^4 y_0 - \frac{3}{128} \Delta^6 y_1 - \frac{3}{256} \Delta^8 y_0 - \text{etc.} \quad (47)$$

$$\sigma_1 - \sigma_0 = \frac{1}{16} (\Delta^4 y_0 + \Delta^4 y_1) - \frac{3}{256} (\Delta^6 y_0 + \Delta^6 y_1) + \text{etc} \quad (48)$$

$$\sigma_1 + \sigma_0 = -\frac{1}{4} \Delta^2 y_1 - \frac{3}{64} \Delta^4 y_1 + \text{etc.} \quad (49)$$

$$\frac{(\sigma_1 - \sigma_0)}{(\sigma_1 + \sigma_0)} = \frac{1}{4} \frac{(\Delta^4 y_0 + \Delta^4 y_1)}{\Delta^2 y_1} - \text{etc.} \quad (50)$$

The Use of Color and Fluorescence Indicators for Determining the Structure of Glasses

W. A. WEVL

Glass Science, The Pennsylvania State College, State College, Pennsylvania

(Received April 25, 1946)

The fact that light absorption and emission depend on the environment of the ions and molecules which absorb or emit light can be used to draw conclusions about the structure of liquids, glasses, and crystals. This method supplements, to a certain extent, information gained from other sources, such as x-ray diffraction data. The method is illustrated by several examples, like the influence of thermal history on the structure of glass and the constitution of the yellow fluorescing zinc orthosilicate.

1. INTRODUCTION

FOR a long time chemists avoided studying macromolecular substances. The classical period of organic chemistry with its brilliant synthesis of natural and artificial dyes of alkaloids and pharmaceutical compounds left the field of high polymers practically untouched. The chemist of this period neglected all those reaction products which had the slightest resemblance to a resin. Reactions leading to an amorphous product which could not be purified by crystallization were considered failures.

This hostile attitude to everything with a high molecular weight changed only when the physicists began to cooperate with the chemist

and when physical measurements became routine methods in the chemical laboratory. With the increase of the molecular weight certain physical properties become more characteristic, whereas the strictly chemical features of the macromolecules lose some of their significance. It becomes difficult to distinguish similar macromolecules of different degrees of polymerization by purely chemical reactions. The distinguishing features are their diffusion speed, viscosity, cohesive forces, and their relaxation times. This is true for both organic and inorganic compounds. The work which shall be discussed here was chiefly designed to study inorganic high polymers, especially silicate glasses.

The vitreous state of matter has certain characteristics which are brought about by the high degree of polymerization and by the fact that the relaxation time of glasses is too high to allow the establishment of an internal thermodynamical equilibrium. Textbooks usually emphasize that glasses are supercooled melts which are instable with respect to their crystalline phases. This is no doubt true for many cases and it explains why certain glasses tend to devitrify, that is, to form crystals. For the understanding of the typical glass properties, however, it is much more important to realize that glasses possess an atomic configuration which had been stable at a certain higher temperature and which on its way to room temperature has been "frozen in" because of the high relaxation time of the glass forming units.

Water can be cooled below its freezing point. The water is then a supercooled liquid but not a glass, it is instable with respect to its crystalline form, the ice, but it is in internal equilibrium. No matter how the supercooled water was prepared, its properties are well defined if the variables, temperature and pressure, are fixed. For glasses, however, this is not true. Depending on its thermal history, the properties of a glass of given composition may vary widely. The density, refractive index, chemical resistivity, or the electric conductivity of a glass is no longer defined by its composition, temperature, and pressure.

The thermal expansion of a supercooled liquid is uniquely defined by a curve which gives the

volume as a function of the temperature. If we allow a borosilicate glass to cool very slowly to room temperature and then determine its thermal expansion, we find a curve like that in Fig. 1a. If the same glass was chilled rapidly, so that the polymerization is incomplete, it shows an entirely different thermal expansion. As indicated in Fig. 1b the expansion at low temperature changes into a contraction around 350°C. This contraction is caused by a polymerization process

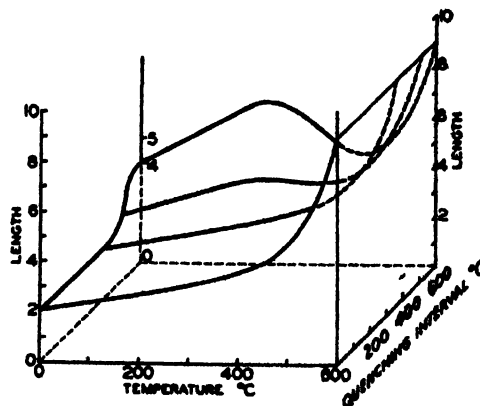


FIG. 2. Space diagram, showing the relationship between volume, temperature, and previous heat treatment.

or by a change in the geometrical arrangement of the glass forming units which could not take place on cooling because of the high relaxation time and the speed of passing through this temperature region.

In order to describe fully how a glass of given composition will change its volume on heating at a certain rate, we need a space diagram which represents the volume as a function of both temperature and previous heat treatment (Fig. 2).

Based on economic and technological considerations such as the source of raw materials, the methods of manufacture, and the processes used for forming the material into its final shape, one has repeatedly tried to draw a sharp demarcation line between inorganic glasses and amorphous organic materials. Modern developments such as glasses free of silica (phosphate glasses) or even free of oxides (beryllium fluoride) and especially the organic derivatives of silica make a sharp distinction very difficult. From a scientific point of view there was never a justification for a distinction between organic and inorganic

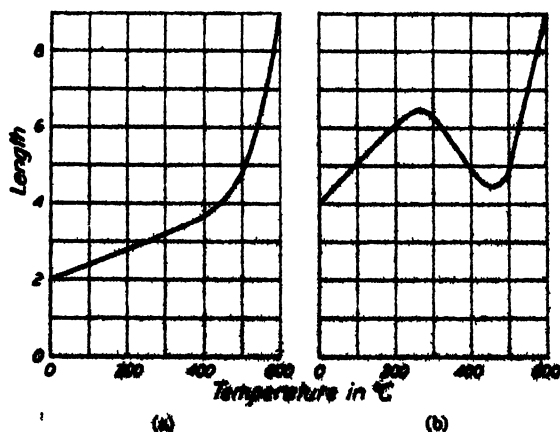


FIG. 1. Schematic curve of the thermal expansion of (a) an annealed glass and (b) a rapidly chilled glass.

glasses. The vitreous state with its characteristics such as optical isotropy, typical thermal expansion and its dependence of all properties on the previous thermal treatment is found among elements, oxides, sulfides, salts, and a great variety of organic compounds. G. Tammann¹ to whom we owe much of the early work on glasses and their properties studied such compounds as brucin, salicin, betol, piperin as well as borates, phosphates, and elemental selenium. These glasses can be studied within a more convenient temperature range and their principal features are not different from those of the high melting silicate glasses.

Just to give you another example of the close analogy of inorganic and organic high polymers, I want to mention that recently we synthesized a purely inorganic material with rubber elastic properties, a hydrated alkali-manganese-phosphate. This was accomplished by cross-linking the long chain-like sodium-potassium-polymetaphosphate molecules by means of a two-valent ion, in other words, by means of a unit having two functional groups. Sodium-potassium metaphosphates can be obtained as high polymers with molecular weights of 100,000. Their aqueous solution forms a viscous liquid. If this solution is allowed to react with divalent ions, in particular with manganese, a complex alkali-manganese phosphate is precipitated. The divalent ions link together the long chains thus depriving them of their free mobility. The resulting rubber-like product shows birefringence when stretched.

The polymerization of P_2O_5 can still be studied by similar methods as that of organic compounds. Despite its tendency to form high polymers, P_2O_5 is soluble to a certain extent in such solvents as chloroform. SiO_2 glass as compared with P_2O_5 glass, however, is insoluble in all solvents. The high viscosity of fused silica indicates that even at 1800°C this compound is still highly polymerized.

The methods for studying the structure of vitreous silica and silicates are greatly limited by the insolubility of these high polymers and by the high temperatures required to decrease their relaxation times sufficiently to bring them within experimental range. So far the Fourier

analysis of the x-ray diffraction patterns has proved to be the most powerful tool for studying the structure of simple glasses. If the number of the atomic species does not exceed 3 or 4, one may obtain fairly reliable data concerning the average interatomic distances and the coordination number of the cations.

B. E. Warren² and his students have perfected this method and were able to derive the atomic structures of several simple glasses such as SiO_2 , BeF_2 , B_2O_3 , and alkali borates and silicates. It must be considered as one of the major results of their work that they found the laws of crystal chemistry to hold true for the vitreous silicates as well as for the crystalline ones.

Before it became possible to interpret the x-ray diffraction patterns of amorphous substances, the author³ and his collaborators used color and fluorescence indicators to obtain insight into their atomic structures. This method is based on the introduction of certain ions of the transition elements into the glass. Their light absorption and fluorescence is affected by the surrounding ions. Additions of new ions, substitutions, and changes in the structure due to temperature and heat treatment reflect on the indicator ions which in turn signal this change of their environment by changing their color or fluorescence.

II. THE USE OF COLOR AND FLUORESCENCE INDICATORS

Changes of color have been widely used to indicate changes in chemical and physical conditions. The acid-base indicators such as methyl orange and phenolphthalein are well known. Cobaltous chloride has been used to indicate the humidity of the air. Since the use of silver mercuric iodide to indicate temperatures above 35°C, by its transition from the red to a yellow modification, many color indicators have been developed which permit us to estimate temperatures of moving machine parts or furnace walls over a wide range.

² B. E. Warren, *J. App. Phys.* 8, 645 (1937).

³ W. A. Weyl, *Veröffentl. Kaiser Wilhem-Institut für Silikatforschung* 7, 167-203 (1935). A rather complete bibliography on the subject of color and fluorescence indicators can be found in the author's monograph on *Colored Glass* prepared for the British Society of Glass Technology. Part I. The Constitution of Colored Glass, *J. Soc. Glass Tech.* 27, 133-206 (1943).

¹ G. Tammann, (a) *Kristallisieren und Schmelzen* (Leipzig, 1903); (b) *Der Glasezustand* (Leipzig, 1933).

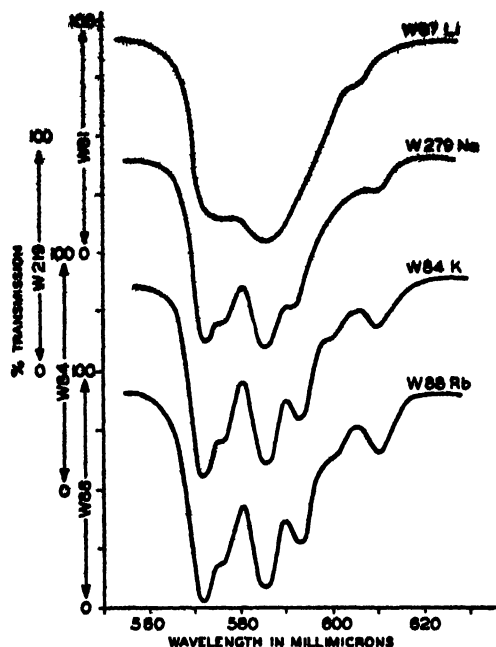


FIG. 3. Light transmission of four alkali silicate glasses containing neodymium as a colorant.

The polymerization of anions to iso- or heteropoly acids is associated with a change in light absorption. Studies in the ultraviolet and in the visible region revealed that the formation of polymer tungstates, vanadates, or chromates causes the absorption band to shift to longer wavelengths. For example, the addition of sulphuric acid to a neutral chromate causes its yellow color of the CrO_4^{2-} to change into the orange of the $\text{Cr}_2\text{O}_7^{2-}$ and finally into the deep red of the $\text{Cr}_2\text{O}_{10}^{4-}$.

An analogous shift of the absorption edge with increasing molecular weight has been observed with alkali polysulfides. Both cases, chromates and polysulfides, have been studied in aqueous solutions as well as in silicate glasses.

F. Weidert was the first to introduce color indicators into the glass melt. When studying the light absorption of glasses containing the rare earth element neodymium, he found that its spectrum is affected by the amount and the nature of the alkali present in the glass. Figure 3 gives the light transmission between 550 and 650 μ for four alkali silicate glasses containing neodymium. The fine structure of the neodymium band is most pronounced in the rubidium glass. Replacing the large rubidium ion by

other monovalent ions of decreasing size causes the fine structure to disappear. The small lithium ion causes the strongest perturbation. The disturbance caused by the electric fields of the alkali ions expresses itself in many different ways. The stronger the electric field of the alkali, the more fluid is the melt and the greater is the tendency of the glass to devitrify. Some potassium silicate glasses require many months of heat treatment to produce crystals. Lithium silicate glasses are so unstable that only small quantities can be obtained without crystallization. Sodium silicates assume an intermediate position. Weidert showed that for a large number of glass compositions the fine structure of the neodymium band can be used as a means to determine the stability of a new glass composition. Blurring of the fine structure was found to be an indication of pronounced tendency to devitrify.

The light absorption of neodymium is caused by electron transitions taking place in rather well-protected inner orbits. The electrons which are responsible for the color are not identical with those which are involved in the formation of chemical compounds. In contrast to the colored salts of the elements of the transition group, all compounds of neodymium have nearly the same pink color. The oxide, the hydroxide, the sulphate, and the ammoniates have nearly identical absorption bands. Only minor shifts in the maxima and in the fine structure indicate that the deformation of the orbits of the outer valence electrons affects to a certain extent the incomplete inner shells.

The author and his student used preferably those color indicators whose light absorption is caused by the outer or valence electrons.

The light absorption of the salts of the elements of the transition group such as Mn, Fe, Ni, Co, Cu, Cr, or V is caused by electron transitions prohibited in the free ions but made possible by a deformation caused by surrounding ions (Molecular Stark effect). The maximum molar extinction coefficient of these ions is therefore usually low and of the order of 0.1 to 1.0. Under certain conditions these ions form complexes with much stronger light absorption which can be attributed to electron transfer bands.

There is no fundamental difference between

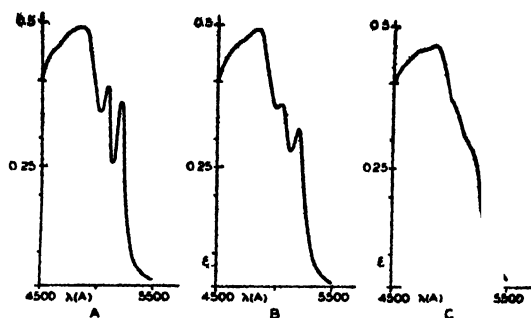


FIG. 4. Light absorption of quinizarin in (a) hexane, (b) hexane-nitrobenzene (7 percent), and (c) hexane-nitrobenzene (24 percent).

the light absorption of these ions, in aqueous solutions and in glasses. The rigidity of glasses, however, prevents their frequent collisions with their neighbors and as a result enables many of them to fluoresce. None of the simple ions such as Mn^{++} , Cu^{+} , V^{++} fluoresces when dissolved in water, but they do so when introduced into glasses. The fluorescence—like the absorption spectrum—is also influenced by the environment as will be discussed later for the case of the divalent manganese ion.

III. THE SPECTRUM OF INDICATORS AS AFFECTED BY SOLVATION

The strength of the forces which different solvents exert upon the solute can vary very widely. They are weak if the solvent consists of molecules which have no permanent dipole and low polarizability. Such is the case for hexane and other "inert" solvents. Molecules with a greater dipole moment and greater polarizability exert stronger forces. These forces can be compared with external electric fields which are asymmetrical and which constantly change their direction with the thermal motion of the molecules. The splitting of spectral lines in external electric fields is well known. In case of asymmetrical fields a blurring of the fine structure results rather than a sharp splitting of spectral lines into their components. Change in direction of the field, as caused by the thermal motion of the surrounding molecules, acts in a similar way.

Many organic compounds which are soluble in non-polar hexane have an absorption spectrum which exhibits some fine structure. Figure 4a gives the light absorption of quinizarin in hexane. The two maxima are most pronounced in non-

polar solvents. Addition of 7 percent of the strongly polar nitrobenzene (Fig. 4b) blurs the sharp maxima and in a mixture of hexane and 24 percent nitrobenzene they have completely disappeared (Fig. 4c).

The light absorption of quinizarin has here been used as an indicator for the polarity of the nitrobenzene. The electric fields of the nitrobenzene molecules exert an influence which can be compared with that of the lithium in neodymium glasses (Fig. 3).

The absorption spectra of many crystalline compounds such as chromium salts resemble those of glasses containing the same color centers, but they differ in their behavior on cooling. The broad absorption bands of the green chromium glasses do not change much in their appearance, if the glass is cooled to the temperature of liquid air. The crystalline chromium salts in this temperature range possess fine structure resembling line spectra. Why do the spectra of crystals and of solutions or glasses respond to temperature changes in a different way? In both cases the transitions of electrons in the chromium ion are affected by the external electric fields of the surrounding ions. The lack of symmetry of these fields in crystals is caused by thermal vibrations. As the temperature approaches absolute zero the intensity of the vibrations decreases and the color centers find themselves in electric fields of identical symmetry which now do not change any more with time. In glasses the alternating electric fields produced by the surrounding ions in thermal motion become also more constant on cooling, but the glass structure causes each color center to be in a slightly different environment. As the light absorption integrates over a multitude of color centers, a blurred spectrum results even if the thermal motion has ceased.

This behavior excludes all those pictures of the glassy state which are based on a heapwork of submicroscopic crystals. The state of disorder must be in molecular dimensions.

Besides its influence upon the fine structure, the solvent determines the exact position of the light absorption. Increased solvation often shifts absorption to shorter waves, indicating an additional energy requirement for the electronic jump. Perhaps the best known example is the

color⁷ of elemental iodine in various solvents. Hexane as an inert solvent allows the iodine molecules to display optical properties similar to those in the vapor state. Solutions of this type are purple, like iodine vapor, and are characterized by a strong absorption band in the green region of the spectrum. Addition of molecules of higher polarity or greater polarizability to the hexane solution causes this band to broaden slightly and to shift to shorter wavelength (Fig. 5). The electronic transition which is responsible for the light absorption in the green region requires now a greater energy. For a theoretical interpretation of this shift we might use the same picture which D. M. Bose and S. Datta developed for explaining the different position of absorption bands and the variation of the magnetic susceptibility of an ion in different lattices. They assumed that a certain interpenetration of the outer electrons of the paramagnetic ion and its environment has to be disentangled before the electron transition can take place. The energy required to loosen the "1-coupling" reflects on the position of the absorption band.

Changing the solvent from hexane to benzene shifts the maximum absorption of iodine from 525 $m\mu$ to 500 $m\mu$. Using toluene instead of benzene causes a similar but greater shift of the absorption band to 485 $m\mu$. The iodine can be used as an indicator for the solvation forces exerted by benzene, toluene, or similar compounds. The further the band is shifted to shorter waves, the stronger is the interaction between solvent and iodine, probably because of greater mutual deformation.

If the solute can be excited to fluoresce, we find a similar influence of the solvent on its light emission, both on the fine structure and on the position of the bands.

Addition of nitrobenzene to a solution of a fluorescing organic compound in hexane quenches the fluorescence. Even the strong fluorescence of mineral oil can be decreased to a small fraction of its original intensity if a few drops of nitrobenzene are added. Under standardized conditions, using a solution of nitrobenzene in hexane, one can use this effect as the basis for working out quantitative methods for fluorometric analysis.

In glasses elemental selenium produces a pink color; its absorption band broadens if potassium

is replaced by sodium or lithium. The same change of composition causes the red fluorescence of the selenium to become fainter and to disappear.*

More spectacular than the quenching of fluorescence in glasses and solutions is the shift of the emission spectrum.

In the case of iodine we have noticed that increased solvation calls for larger light quanta, thus shifting the absorption to shorter waves. As far as the fluorescence is concerned the increased solvation expresses itself in the emission of smaller light quanta and consequently in a shift of the fluorescence color to longer waves. Using uranium glasses as an example, one finds that the uranyl group can produce a group of five relatively sharp emission bands in the blue-green part of the spectrum (potash glasses, certain phosphates, and borates) and as the other extreme a broad emission band with little or no fine structure ranging from green to orange (lithium or high sodium silicates).

The organic chemistry offers numerous examples for the variation of the fluorescence color

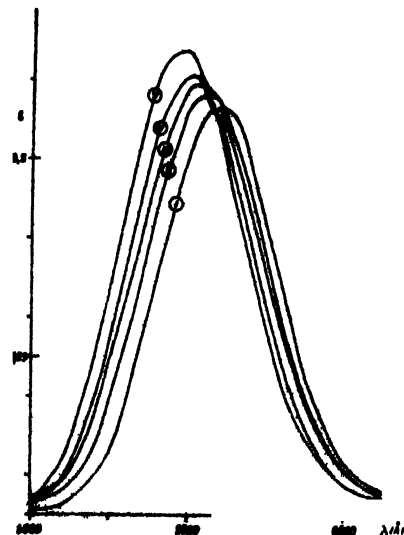


FIG. 5. Light absorption of iodine in hexane-benzol.

Curve 1: Hexane pure.
Curve 2: Hexane 14 percent benzol.
Curve 3: Hexane 28 percent benzol.
Curve 4: Hexane 54 percent benzol.
Curve 5: Benzol pure.

* The fluorescence of cations in silicate or phosphate glasses can be improved if the O^{--} ions are partly replaced by F^- ions. Introducing the more deformable Br^- and I^- ions—even in small amounts—on the expense of the less polarizable O^{--} quenches the fluorescence.

TABLE I.

Solvent	Diaminoterephthalic acid ester para	
Hexane	green	blue
Benzene	green-yellow	green-blue
Ether	green-yellow	green-blue
Chloroform	green-yellow	green
Acetone	yellow	green
Pyridine	orange-yellow	green
Ethyl alcohol	orange	yellow-green

with the nature of the solvent. The most striking shift has been reported for the dimethylnaphth-eurhodine. In hexane this compound produces a strong bluish green fluorescence. With increasing polarizability and dipole moment of the solvent, the fluorescence is shifted to green, yellow, orange and finally becomes red in those solvents which contain strong polar groups, such as the alcohols. One is easily tempted to give a more specific interpretation of this phenomenon and correlate the maximum of the emission bands with the dielectric constant or the dipole moment of the solvent. One should keep in mind, however, that the deformation or interpenetration of the electrons responsible for the change of light absorption and emission depends upon a number of factors among which the permanent and induced dipole moments, the polarizability, size, shape, and the position of the responsible atomic group within the molecule are the most important.

The interaction between solvent and solute can be intensified by introducing polarizable groups into the solvent, as can be seen from the influence of benzene and toluene on the light absorption of iodine. It can also be changed by introducing or varying the polar groups of the solute.

Table I describes the influence which the solvent has on the fluorescence of the esters of para- and meta-diaminoterephthalic acids. Table II describes the change of fluorescence of the para-diaminoterephthalic acid ester dissolved in ether* and in alcohol if the amino groups are made inert by acetyl groups and partly or fully replaced by hydroxyl groups.

Increasing the forces acting between a solute and its solvent shifts the light absorption to shorter and the light emission to longer wavelengths, no matter how this increase in the mutual deformation and attraction is accom-

plished, by modifying the solvent or by modifying the solute.

2. The Absorption Spectrum as Affected by Chemical Changes

We have interpreted the loss of fine structure and the gradual shift of the absorption bands to shorter wave-length as the result of increased solvation. Solvation must be pictured dynamically. The solute molecules attract and orient surrounding solvent molecules, but thermal motion does not permit the formation of more permanent configurations. At low temperature especially if the forces acting between solvent and solute are strong, one may expect associations of longer lifetime to form between solvent and solute. In this case one speaks of compound formation. Iodine for example forms three groups of solutions. In hexane or carbon tetrachloride it produces the same purple as the vapor and one assumes little interference between the I_2 molecules and the solvent. For benzene and toluene we have found some interaction which causes the absorption band to be slightly shifted: these solutions are red. In alcohol, ether, or in an aqueous solution of potassium iodide, iodine dissolves with a deep brown color, and it is generally assumed that a more or less defined chemical compound has formed. In the potassium iodide solution it is the anion $(I_3)I^-$. Such an association of longer lifetime than the previously pictured dynamic solvation can give rise to different electron transitions.*

Figure 6 indicates the change of the light absorption of iodine in hexane upon addition of ether. The formation of a compound between ether and iodine is indicated by the appearance

TABLE II.

Derivative	Ether solution	Alcoholic solution
2,5-diamino ester	yellow	orange-yellow
2-amino-5-oxyester	green	yellow-green
2/5-dioxyester	blue	blue
Monoacetyldiamino ester	blue-green	green
Diacytyldiamino ester	violet-blue	violet-blue

* Whether a purple, red, or brown solution is formed depends not only on the nature of the solvent but also on the temperature. Red solutions become purple on heating because of decreased solvation. Brown solutions turn red because of the thermal dissociation of the labile compounds.

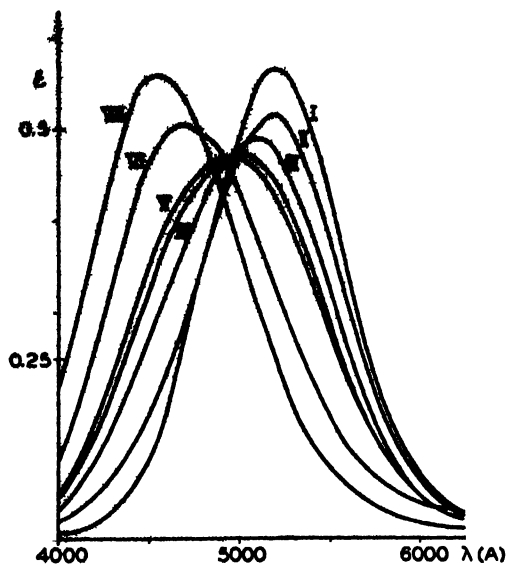


FIG. 6. Light absorption of iodine in hexane-ether.

Curve 1: Hexane pure.
Curve 2: Hexane 2.28 percent ether.
Curve 3: Hexane 4.56 percent ether.
Curve 4: Hexane 7.15 percent ether.
Curve 5: Hexane 8.60 percent ether.
Curve 6: Hexane 28.6 percent ether.
Curve 7: Ether pure.

of a new absorption band. The position of the new band depends on the nature of the reactant. It is 450 $m\mu$ for ether and 440 $m\mu$ for alcohol. In contrast to the addition of benzene to the iodine hexane solution, we observe here the formation of a new color center on the expense of the original one which we attributed to free, but solvated, I_2 molecules.

This type of change, appearance of a new spectrum and simultaneous disappearance of the original one, is characteristic for all chemical changes. It is widely used in colorimetry and in the field of glasses it has given us much valuable information about their chemical reactions. As an example for an oxidation-reduction process in glass, we give Fig. 7 demonstrating the reduction of CrO_3 to Cr^{++} by means of As_2O_3 . If chromium compounds are added to the glass batch, a green glass is formed. The color is produced by both Cr^{++} and Cr^{+++} which are in equilibrium. The position of the equilibrium depends upon the melting temperature, furnace atmosphere, composition of the base glass and additions such as arsenic. The Cr^{++} has its main absorption in the ultraviolet and in the short wave region of the visible. It does not absorb in the orange, where

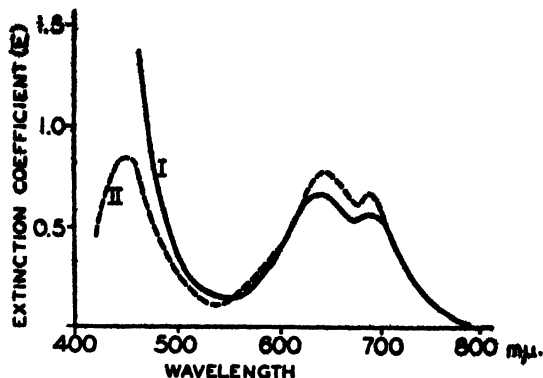


FIG. 7. Light absorption of chromium containing glass, melted under oxidizing and under reducing conditions.

the Cr^{++} has an absorption maximum. The height of this band can, therefore, be used to evaluate the amount of Cr^{++} and with it changes of the oxidation-reduction equilibrium.

These chemical reactions are of great importance to the glass technologist, especially to the one who has to make reproducible color filters. From the viewpoint of the glass structure another type of compound formation is far more significant, a change which cannot be expressed by the conventional chemical equations.

3. The Spectrum as an Indicator of Changes in Coordination

The different colors produced by iron, manganese, chromium, or vanadium result chiefly from these elements being capable of forming a number of different ions in glass. Iron occurs as Fe^{3+} and Fe^{2+} ; vanadium as V^{5+} , V^{4+} , and V^{3+} ; manganese as Mn^{2+} and Mn^{3+} . Each ion is an individual and its optical properties are just as different as its chemical ones.

Other elements such as cobalt and nickel when introduced into glasses can form only one type of ion, namely Co^{2+} and Ni^{2+} . Nevertheless there are cobalt glasses which are pink and others which are deep blue. Nickel-containing glasses range from yellow over grey shades to deep purple. In these cases the different absorption spectra are caused by different coordination complexes in which the Co^{2+} or Ni^{2+} participate as the central cation. This is known from the behavior of these ions in aqueous solutions and in various host lattices.

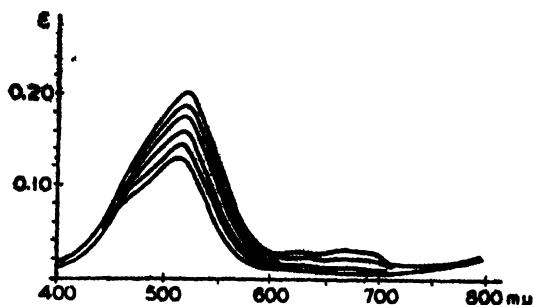


FIG. 8. Light absorption of the Co^{++} ion in aqueous solutions containing increasing amounts of calcium nitrate.

Cobalt ions diluted in aqueous solutions are pink. They have an absorption band in the green part of the spectrum. Color centers are the completely hydrated Co^{++} ions. Dehydration caused by increasing the temperature, addition of dehydrating compounds such as calcium nitrate or glycerol produces a slight change in the light absorption. The maximum is shifted to longer wave-length and its height increases. Dehydration produces the opposite effect of increased solvation. The electronic jump becomes more probable and is caused by light quanta of smaller size. Figure 8 represents the deepening of the pink of a diluted cobaltous nitrate solution upon addition of calcium nitrate.

The addition of halides such as LiCl or CaCl_2 in its first stages produces the same effect. If the halide concentration exceeds a certain limit, one observes the formation of a deep blue complex ion such as $(\text{CoCl}_4)^{2-}$. This process is pictured in Fig. 9.

Whenever an anionic complex CoX_4 is formed, one obtains an intensive light absorption in the long wave part of the visible spectrum. If X is Cl^- , CNS^- , or O^{2-} , the color is deep blue. Using more polarizable ions such as I^- causes the bands to shift to longer wave-length and green complexes result. This has been observed in solutions as well as in borate glasses containing iodine. A blue pigment is obtained by introducing cobalt into a crystal lattice where it is surrounded by four oxygens. That is the case in the cobalt aluminate Al_2CoO_4 (Thenards blue). Mixed crystals between MgO and CoO on the other hand are pink because they have NaCl structure and each Mg^{++} and Co^{++} is surrounded by six oxygens. This simple rule that CoO_6 complexes

are pink and CoO_4 complexes are blue holds true only for those crystals which consist of rare gas ions. Ions with incomplete outer electrons exert a strongly deforming influence upon the oxygen ions which in turn reflects upon the $\text{Co}-\text{O}$ bond. Cobalt zincate for example is green.

The same two types of color centers can be formed in glasses. CoO_4 groups can be considered complex anions comparable to the SiO_4 groups. It seems reasonable to assume that in the random three-dimensional network of SiO_4 tetrahedra an occasional SiO_4 group can be replaced by a CoO_4 group. In this case the cobalt ion plays the same role as the silicon. According to B. E. Warren ions with these functions are called "network forming" ions. They usually have the coordination number four; only boron in the B_2O_3 glass has the coordination three. The polymerization of these randomly arranged XO_4 groups leads to a huge anionic three-dimensional network in whose interstices the "network modifying" cations are distributed.

The modern concept of N.W.F. and N.W.M. cations corresponds roughly to the older picture of acid and basic glass forming oxides. Only the acids such as SiO_2 , P_2O_5 , B_2O_3 can form glasses by themselves. The glass chemist also used the term "amphoteric" for oxides which in some glasses assume the role of an acid and in others

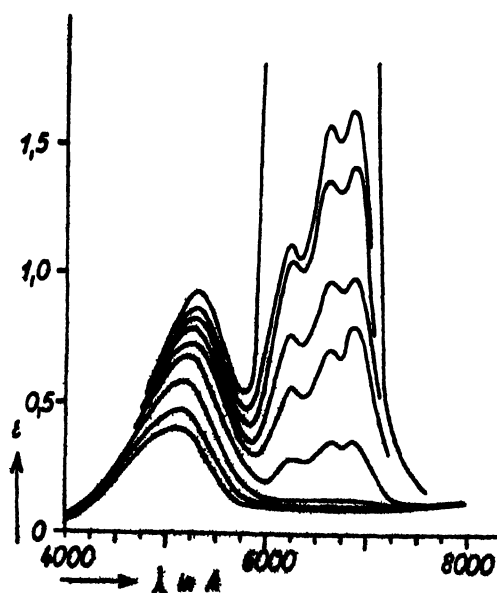


FIG. 9. Light absorption of the Co^{++} ion in aqueous solutions containing increasing amounts of lithium chloride.

that of a base. In modern parlance, Al^{3+} , for example, can act as an N.W.F. cation by entering the glass structure as AlO_4 tetrahedra or remain in interstices where it is surrounded by a larger number of oxygens in greater distance.

The distinction between N.W.F. and N.W.M. cations is based on the different strengths of their electric fields. Small ions of high charge are the first to surround themselves with oxygen ions. Tetrahedral groups are particularly favorable to glass formation. If the chemical composition of the melt does not provide sufficient oxygen ions to permit the formation of independent SiO_4 groups, the tetrahedra have to share oxygens.* This polymerization increases the viscosity of the melt very rapidly and finally leads to a rigid three dimensional network, the glass.

If the chemical composition of the melt provides sufficient oxygens to form independent SiO_4 tetrahedra as in the case of sodium orthosilicate, Na_4SiO_4 , there is no more necessity for sharing the oxygens and with it for polymerization. Such melts are fluid and can crystallize easily on cooling.

Whether or not a melt forms a glass depends, therefore, on the charge and size of the strongest cation and on their number as compared with that of the oxygen. V. M. Goldschmidt has derived certain rules for glass formation applicable to oxide and fluoride glasses, which were based on the probable coordination, because of size and charge of the cations and the cation to oxygen ratio.

The validity of these rules is limited to the anions of the rare gas type. The strongly deforming influence which ions with incomplete outer electronic shells exert upon the large anions leads to bonds which cannot be predicted anymore by simple rules, because the ionic radii of the deformed ions have lost their significance.

The orthosilicates of noble gas ions do not form glasses. However, lead and zinc orthosilicates are known in vitreous condition. We have here a group of compounds which bridge the gap between the strictly ionic and the organic glasses.

* The sharing of oxygens between two SiO_4 tetrahedra can be considered one principle of glass formation. It is the chief reason why silicate melts polymerize.

Whereas size and charge of the cobalt and nickel ions should classify them as N.W.M. cations or basic glass constituents, their incomplete outer electronic shells enable them to form CoO_4 and NiO_4 complexes in equilibrium with

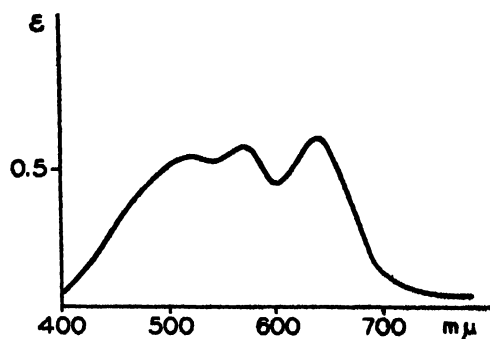


FIG. 10. Light absorption of the Ni^{2+} ion in a potassium silicate glass (purple type).

those Co^{2+} and Ni^{2+} which have their places in interstices or N.W.M. positions. The relative number of XO_4 groups formed by their cations might be very small but they possess very intensive light absorption. The covalent bonds between Co^{2+} and O^{2-} in the strongly bonded CoO_4 complexes probably makes possible horizontal electron transitions, producing electron transfer spectra instead of the prohibited vertical transitions of Co^{2+} ion.

Figure 10 shows the absorption spectrum of a purple Ni^{2+} containing silicate glass. We attribute this characteristic absorption to NiO_4 groups. They form preferably in those glass compositions where the Ni^{2+} can successfully compete with the other cations and form independent anionic groups (rubidium silicate). Glasses built up from ions with higher electric potential (lithium silicate or phosphate glasses) prevent the formation of NiO_4 groups and shift the Ni^{2+} into N.W.M. or interstitial positions. In this case the Ni^{2+} has a higher coordination, maybe 6, is farther away from the anions, and its light absorption is different (Fig. 11). Glasses of this type are yellow. The Ni^{2+} ion is thus an indicator for the availability of anions in the glass. Its light absorption is a function of the extent to which it succeeds in forming its own anionic or covalent group. A few results obtained with Ni^{2+} as color indicator shall be mentioned.

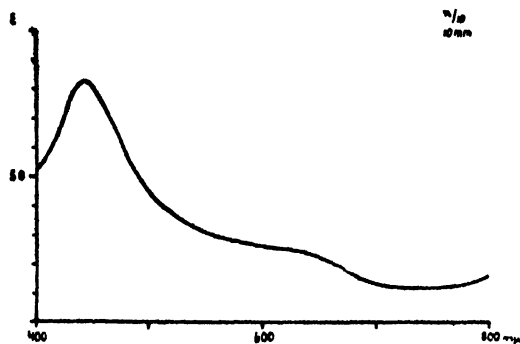


FIG. 11. Light absorption of the Ni^{2+} ion in a lithium-sodium silicate glass (yellow type).

(1.) Sodium silicates containing Ni^{2+} are grey. Their absorption spectra show the presence of both NiO_4 and NiO_6 groups. Replacing Na^+ by the smaller and therefore more effective Li^+ causes the color to change to yellow. Substitution of K^+ , and even more so of Rb^+ , for Na^+ produces purple glasses. The weaker the electric field of the other N.W.M. cations, the greater are the chances for the Ni^{2+} ions to compete for their oxygens and form NiO_4 groups.

(2.) If the gray Ni^{2+} containing sodium silicate glasses are heated into the softening range, their light absorption indicates an increase in the number of NiO_4 groups (Fig. 12). On cooling to room temperature some of the nickel ions change back from N.W.F. to N.W.M. positions. The indicator ion signals a redistribution of the anions with temperature, not merely a change of the interionic distances. In the crystalline state such a change of geometrical arrangement of the ions produces a new modification. It has to take place abruptly because of the symmetry requirements of the lattice and it is accomplished by cooperative maneuvers. Glasses show changes in their structure which are comparable with modification changes of crystalline silicates but the adjustment takes place gradually; the single units do not have to cooperate to accomplish the change.

(3.) The light absorption of a chilled Ni^{2+} containing sodium silicate glass indicates the presence of more NiO_4 groups than in the annealed condition (Fig. 13). This indicates that glasses of different heat treatment possess different atomic arrangement.

The fact that glasses of different heat treat-

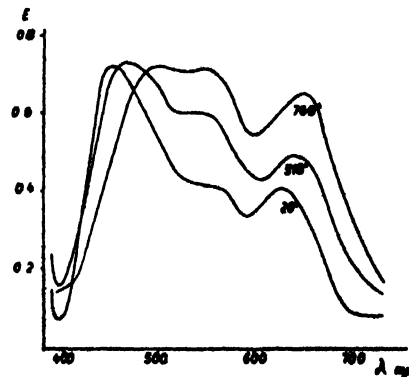


FIG. 12. Light absorption of the Ni^{2+} ion in a sodium silicate glass of different temperatures.

ment possess very different physical and chemical properties can be explained on the same basis as the difference of properties of the high and low temperature form of a certain crystalline compound. Now we can understand the meaning of the volume contraction which the quenched glass exhibits on heating. It corresponds to a change of the high temperature modification into the stable low temperature form. The rapid chilling of the glass had frozen in a high temperature form just as the cooling of molten sulphur produces the rhombic rather than the monoclinic form.

The change in coordination or the role which a cation plays in the glass structure effects the emission spectrum in a very similar way as its absorption. The divalent manganese is one of the best examples. MnO_4 groups produce green and Mn^{2+} in N.W.M. positions produces red

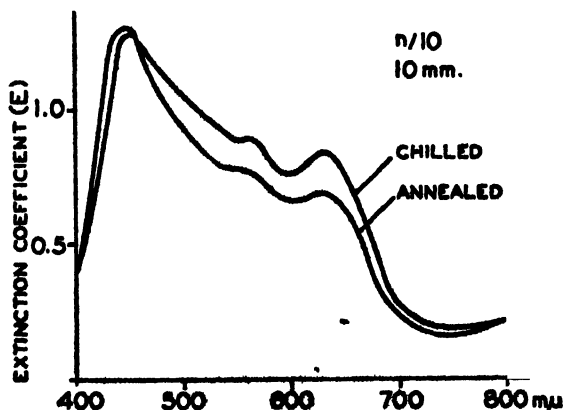


FIG. 13. Light absorption of a nickel-sodium silicate glass in annealed and chilled conditions.

fluorescence. The principles involved in using Mn^{2+} ions as fluorescence indicators are mostly the same as described for Ni^{2+} as color indicator. N. J. Kreidl⁴ has recently published a paper on this subject where he describes the use of the Mn^{2+} for exploring new fields of optical glasses.

It might be interesting, however, to mention that this fluorescence indicator had been used to draw conclusions concerning the structure of the yellow fluorescing zinc silicate. This compound represents an unstable modification of the zinc-

orthosilicate activated by Mn which for a long time could not be obtained in well-developed crystals. It forms when ZnO and SiO_2 are allowed to react at relatively low temperature and by some authors it was considered to be amorphous. Only recently we could prepare this modification in better crystals as a devitrification product of some multicomponent zinc silicate glasses. These glasses precipitate zinc orthosilicate on heating and form the stable green modification at high temperature and according to "Ostwald's Step Low" the yellow unstable modification at lower temperature.

⁴ N. J. Kreidl, J. Opt. Soc. Am. 35, 249-257 (1945).

Calendar of Meetings

July

- 12-13 American Physical Society, Berkeley, California
- 17-19 American Society of Civil Engineers, Spokane, Washington

August

- 14-15 Institute of Aeronautical Sciences, Los Angeles, California
- 19-20 Mathematical Association of America, Ithaca, New York
- 20-23 American Mathematical Society, Ithaca, New York

September

- 9-13 American Chemical Society, Chicago, Illinois
- 16-20 The Instrument Society of America, Pittsburgh, Pennsylvania
- 18-21 Illuminating Engineering Society, Quebec, Canada
- 19-21 American Physical Society, New York, New York
- 30-October 3 American Society of Mechanical Engineers, Boston, Massachusetts

October

- 13- 5 National Electronics Conference, Inc., Chicago, Illinois
- 9-11 American Society of Civil Engineers, Kansas City, Missouri
- 25 American Mathematical Society, New York, New York

November

- 7-9 Conference on Electrical Insulation, Division of Engineering and Industrial Research, National Research Council, Baltimore, Maryland
- 29-30 American Physical Society, Minneapolis, Minnesota

December

- 2-6 American Society of Mechanical Engineers, New York, New York
- 17 Institute of Aeronautical Sciences, New York, New York
- 26-28 Geological Society of America, Chicago, Illinois

Here and There

New Appointments

N. B. Nichols has been appointed director of the research division of the Taylor Instrument Companies. Carlton W. Miller and William A. Mersman, both formerly of the Radiation Laboratory, have joined the research department, and Walter G. Countryman has returned to research work at Taylor after four years of military service.

Radio Corporation of America recently announced two appointments. John L. Reinartz has returned to the company, to be in charge of the amateur radio program in the Commercial Engineering and Power Tube Sections. John F. Rider, radio engineer and writer, has been retained by the Victor Division as a consultant on test equipment.

Fellowship at Michigan State

Michigan State College has announced a \$1000 fellowship available for graduate study and research in physics for one year from September 1946. This fellowship has been established from a Frederick Gardner Cottrell special grant-in-aid, contributed by The Research Corporation of New York, for a study of some of the properties of ions and electrons in gases. Applications will be considered from any able individuals who hold bachelors' or masters' degrees from approved institutions and whose work shows promise of originality. The Fellow will not be required to teach. Inquiries should be addressed to Dr. R. C. Huston, Dean of the Graduate School, Michigan State College, East Lansing, Michigan.

Laboratory Clinics on Polymer Research

The Institute of Polymer Research at the Polytechnic Institute of Brooklyn is sponsoring a series of summer laboratory clinics for 1946. Five sessions will occur, June through September. A fee is charged for each clinic, and attendance for each one will be limited.

Journal of Polymer Science

Journal of Polymer Science is now being published bi-monthly by Interscience Publishers, Inc., and Elsevier Publishing Company, Inc. It is devoted to the advancement of our fundamental knowledge of the physics and chemistry of polymers. The new journal is an outgrowth of, and a successor to, the *Polymer Bulletin*, which has ceased to appear. It reflects the consolidation of independent plans of the Elsevier Publishing Company, Inc., of Amsterdam, and Interscience Publishers, Inc., of New York, to publish separate journals on the same topic. The *Journal* will be international in scope. All papers will be published in the English language.

Duke University Fellowships

Duke University is offering ten special research fellowships in microwave electronics and molecular physics with stipends from \$1600 to \$2400. Graduates of recognized colleges who have majored in physics are eligible. Some of the appointments are open to persons with only

an undergraduate minor in physics, provided they have a major in chemistry or an equivalent training in electronics. Individuals interested in these appointments may write for particulars to the Chairman, Department of Physics, Duke University, Durham, North Carolina.

Conference at Texas A. & M.

Research personnel in industry and college teachers were invited to attend a graduate conference at College Station, Texas, devoted to the behavior of mass spectrometers, electron microscopies, and other electronic devices whose operation depends on the action of electron or ion beams. The conference was conducted by the Electrical Engineering and Physics Department of the Agricultural and Mechanical College of Texas June 24 through July 6. The lecturers were Dr. Ladislav Marton of the Division of Electron Optics of Stanford University and Dr. John A. Hipple of the Westinghouse Research Laboratories.

Film on Compressed Air

"Our Industrial Air Power" is the title of a new educational film on the many industrial uses of compressed air which is available for showing before students, engineering societies, and other groups. Among the sequences in the 25-minute, 16-mm sound-color film are: "Fundamentals of Air Compression," "Properties of Compressed Air," "How Different Types of Compressors Work," and "How Compressed Air Is Used." The film is available without charge on a loan basis. Write Quincy Compressor Company, Quincy, Illinois.

Army Research and Developments Program

Early in May the creation of a Research and Developments Division as a top-level General Staff organization to coordinate Army research with the activities of industry and educational institutions was announced by Secretary of War Robert P. Patterson. At that time General Dwight Eisenhower, Chief of Staff, had completed plans for the operation of the Division on a level with the Personnel Intelligence, Training, Supply, and Operations section of the General Staff.

The Division has primary interest in the application of national scientific resources to the solution of military problems. The director will act as liaison between Army planning chiefs on one hand and industry and the universities and research laboratories on the other. Great civilian advances are foreseen as a result of much of the Army-sponsored research.

Cosmic-Ray Investigation

In search of fuller information in regard to cosmic-ray intensities as they vary with latitude and altitude, the National Geographic Society, the United States Army Air Forces, and the Bartol Research Foundation jointly conducted in May a series of four round-trip flights in a specially equipped B-29 bomber. The principal apparatus carried in the plane consisted of multiple banks of Geiger counters, designed by Dr. W. F. G. Swann, Director of the Bartol Research Foundation.

New Books

Physics of the Twentieth Century

BY PASCUAL JORDAN. Pp. 185+xii. Philosophical Library, New York, 1944. Price \$4.00

This is a simple little book, addressed to the non-physicist who is curious about the results of modern physics, and intended to instruct the philosopher. Its merit lies in its wealth of interesting examples designed to clarify the reputedly paradoxical teachings of relativity theory and quantum mechanics. The difference in outlook between classical mechanics and modern theory is very impressively sketched. Unfortunately, crudities of translation and limitation of scope render it inferior, on the whole, to the elegant expositions of Eddington.

The interpretation of quantum mechanics, to which a large part of the book is devoted, involves an unreserved avowal of the early view which saw in the new doctrine a mathematical reflection of the uncontrollable effects of measurement on physical objects, an establishment of fundamental discontinuity in nature, and a thorough rejection of causality. Very little attention is paid to the formulation, implicit in the work of Dirac and Schrödinger, according to which the concept of "physical states" has undergone modification. As a result, comments on crucial matters are, in the reviewer's opinion, sometimes superficial. Thus when the charge distribution of the electron in a hydrogen atom is asserted to owe its diffuseness to the unpredictable bombardment by gamma-ray quanta coming from the famous microscope, no explanation is given as to why the distribution should not depend on the frequency of the gamma-rays.

As a philosopher, Jordan seems to urge two things: a recognition that materialism has been liquidated, and a bright-eyed positivism willing to accept all empirical facts even if they spell religion. His deliberations dealing with the theory of knowledge do not, it seems, rise to the level of technical competence that would impress philosophers.

In spite of its faults, the book makes pleasant reading. Its misfortune is that it tries to enter a market—that of English scientific literature—in which the standard of excellence for works of this sort is extremely high.

HENRY MARGENAU
Yale University

New Booklets

School of Mineral Industries, The Pennsylvania State College, recently published its Circular 20, *Opportunities for Production Work through Mineral Industries Research*, a 36-page pamphlet generously illustrated with photographs and drawings. Free on request. Address the School at State College, Pennsylvania.

American Optical Company, Scientific Instrument Division, Buffalo 11, New York, recently issued a four-

page folder describing the Spencer Dark Field Quebec Colony Counter which has been placed on the market.

Televiso Products Company, 7466 West Irving Park Road, Chicago 34, Illinois, has published Bulletin No. 31, a four-page description of the Series 200 Vacuum Tube Voltmeter.

Scientific Development Corporation, 614 West 49 Street, New York 19, New York, began in May a new publication service entitled *Public Domain*. Each weekly issue will be a 250-page reference book containing over 1000 patents due to expire four weeks after date of issue, which will then be in the public domain, and open to exploitation by whoever cares to use them. Each patent shown will include a good reproduction of a draftsman's drawing together with a digest of typical claims and salient features. Charter subscriptions are offered for one year at \$45; for six months at \$25; and for 10 weeks at \$10.

Acme Scientific Company, 200 North Laflin Street, Chicago 7, Illinois, has issued its *Bulletin 453*, which gives an extensive listing of stock sizes of optical flats, both single and double surfaced.

Nickel Steel Topics, published at 67 Wall Street, New York 5, New York, discusses in its March, 1946 issue the use of nickel alloys in tractor coil springs, aircraft engines, oil well drilling equipment, cutting saws, and testing machinery.

Precision Equipment Company, 32 North State Street, Chicago 2, Illinois, has published a new catalog for all factory superintendents, maintenance engineers, purchasing agents, and other plant executives interested in industrial equipment and supplies. Any items of industrial equipment or supplies not listed in this catalog will be procured by Precision's Industrial Service Department.

The Tin Research Institute, Frascri Road, Greenford, Middlesex, England, has issued a 32-page report of its work covering the years 1942-1944. Free on request. Technical representation for the Institute in the United States is furnished by the Battelle Memorial Institute, 505 King Avenue, Columbus, Ohio.

Castle Films, Inc., 30 Rockefeller Plaza, New York, New York, is offering free to users of training films a new catalog listing for the first time all U. S. Government 16 mm sound films and 35 mm film strips, totaling 1158. They include training and other types of educational films produced by the Navy, War Department, Department of Agriculture, and U. S. Public Health Service.

The March and April *Ohmite News*, published by Ohmite Manufacturing Company, 4835-41 Flournoy Street, Chicago 44, Illinois, featured a biographical sketch of Sir Joseph John Thomson, discoverer of the electron, and a description of the new Ohmite Laboratory for Precision

Electrical Measurements at Illinois Institute of Technology. Two pages. Free on request.

The Frontier is published quarterly by Armour Research Foundation of Illinois Institute of Technology, Chicago 16, Illinois. The March 1946 issue features articles on "Limitations of Hardenability Indices," "Recent Developments in the Use of Mold Bran," "Electron Microscopy," and "Drinking Water Made Palatable by Chlorine Dioxide." Twenty pages, illustrated.

The General Radio Experimenter for April 1946, published by General Radio Company, 275 Massachusetts Avenue, Boston 39, Massachusetts, describes an electronic sequence timer which permits multiple exposures on a single film to be made with the Microflash. Eight pages, illustrated.

The Main Library of General Electric Company, Schenectady, New York, issues monthly an eight-page review of recent periodicals, giving the title, source, and sometimes a very brief abstract, of current technical articles. The leaflet is called *Library Service, College Edition*.

A twenty-page pamphlet is available entitled *Industrial Research Progress at Armour Research Foundation of Illinois Institute of Technology, 1944-45*, describing its facilities and activities in physics research, as well as those in the other major fields of chemistry, metallurgy, engineering, etc. The Foundation is a non-profit institution established in 1936 to render a research and experimental engineering service to industry. Copies may be obtained by addressing Armour Research Foundation, Chicago 16, Illinois.

Leeds and Northrup Company, 4934 Stenton Avenue, Philadelphia 44, Pennsylvania, has issued Catalog ED,

Galvanometers and Dynamometers. This is a new, completely rewritten edition of earlier Leeds and Northrup galvanometer catalogs. 37 pages. Free on request.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, has issued Bulletin 154-64, which describes Gaertner spectroscopes of the Bunsem-Kirchhoff type. Four pages. Free on request. The same company has also published Bulletin 153-64 on automatic linear dividing machines. Twenty pages. Free if requested, on company letterhead.

North American Philips Company, Inc., 100 East 42d Street, New York 17, New York, has released a new four-page booklet (R1022) entitled "Geiger-Counter Used in Powder Metallurgy." Free on request.

Electro-Voice, Inc., 1239 South Bend Avenue, South Bend 24, Indiana, has issued an illustrated four-page bulletin on its new Model 950 Cardax, a cardioid unidirectional crystal microphone. Free on request.

Interchemical Review, winter issue of 1945, published by Interchemical Corporation, 432 West 45th Street, New York 19, New York, has the following table of contents:

Survey of Sorption Chemistry	EDMUND N. HARVEY, JR.
Our Part in the War Effort	JOHN O'CONNOR
A Tour of Bombed Nazi War Plants	ERNEST W. PITTMAN

Bausch and Lomb Optical Company, Rochester 2, New York, in its spring 1946 issue of *The Educational Focus* presents the following articles:

Optics in Scientific Crime Detection	J. EDGAR HOOVER
Color Photomicrography	J. V. BUTTERFIELD
Uses of Optical Equipment in the Aluminum Industry	STANLEY V. MALCUIT

Journal of Applied Physics

Volume 17, Number 8

August, 1946

Oxide Coated Cathode Literature, 1940-1945

JOHN P. BLEWETT

Research Laboratory, General Electric Company, Schenectady, New York

1. INTRODUCTION

BEFORE the war, it appeared that the mechanism of the oxide coated cathode was becoming fairly clear. The impurity semiconductor picture developed by the authorities on the theory of the solid state seemed to be a good representation of the behavior of an oxide coated cathode. Many phenomena which were well authenticated experimentally and which had been mysteries as far as interpretation was concerned, dropped neatly into place and numerical agreement was achieved between experiment and theory within the rather wide limits of experimental error. It began to appear that sufficiently careful control over such parameters as particle size, packing density, and so forth would lead to accurately reproducible results predictable from theory.

The situation fifteen years ago was somewhat different. At that time it was felt by such authorities as Reimann, Lowry, and others that the interface between oxide and core metal played an important part in the oxide cathode mechanism. Proponents of this view were silenced by the work of Becker and Sears and others around 1931, who seemed to have demonstrated that the outer surface of the oxide and the body of the oxide were the only regions which required consideration.

In the light of this history, it is particularly interesting that experiments performed during the war and reported elsewhere in this issue,

lead not only to reconsideration of the oxide-core interface, but also to a possible re-evaluation of the whole oxide cathode theory. The high pulsed currents which have been drawn from oxide cathodes during the war must have resulted in conditions within the cathode which were far removed from equilibrium. It therefore becomes increasingly doubtful that correct results can be deduced from statistical mechanics which is a discipline related only to matter in equilibrium. The late Professor Fowler in the treatment of semiconductors in his monograph, *Statistical Mechanics*, emphasized this fact as follows: "It should be remembered that all these formulae are only valid so long as the rate of emission of electrons is very slow compared with the rate of readjustment of the internal equilibrium. The process creating the free electrons and free holes must be able to maintain the normal equilibrium supply. This is an essential assumption and there is no simple means of investigating its validity—nor any *a priori* reason why it should be true."

Mr. Hahn of our laboratory has suggested that future theoretical work should include the concept of mass motion and has pointed out that the basic mathematical techniques are already available in Chapman and Cowling's book¹ on the dynamics of gases.

¹S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, New York, 1939).

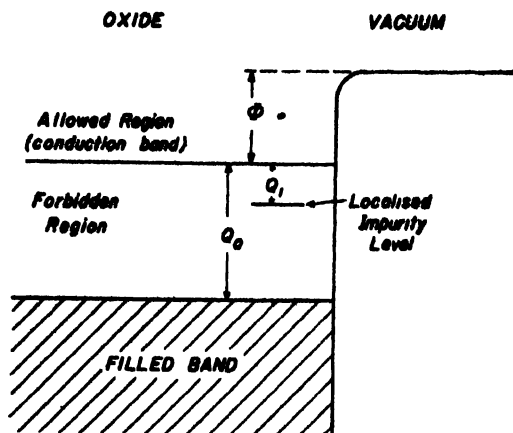


FIG. 1. Energy level distribution for an impurity semiconductor.

Most of the recent work on oxide coated cathodes has been performed under the veil of secrecy imposed by necessary wartime restrictions. A literature survey will thus be, at best, of secondary interest. Several important papers have appeared in the literature, however, and are worthy of mention by way of introduction to the survey of wartime research presented in this issue of the *Journal of Applied Physics*. The bibliography presented herewith is intended to be a supplement to the bibliography included in the author's 1939 review paper.

2. PROPERTIES OF BARIUM

The disagreement between Becker and Moore and Benjamin and Jenkins as to whether or not barium migrates over the surface of tungsten was resolved in 1940 by Becker and Moore who obtained one of Benjamin and Jenkins' experimental tubes and showed that Benjamin and Jenkins' tungsten surface had been slightly contaminated with oxygen. When this contamination was removed, Becker and Moore demonstrated that barium migration does take place in the temperature range above 970°K. They showed also that the migration is predominantly toward the negative end of the filament, so that the barium must have been at least partly ionized.

Timofeev and Aranovich (1940) studied the secondary emission from O-Ba and O-Mg surfaces and pointed out that the stability of such surfaces up to temperatures of the order

of 600°C makes them interesting for practical secondary emitter applications.

Brüning (1940) pointed out that an alloy of copper and barium, when coated on tungsten, possesses a temperature stability much greater than that of a coating of pure barium on tungsten. His experiments indicated that barium evaporates from such a surface at 1300°K at a rate slower than the rate of evaporation of barium from tungsten at 1000°K. Since the alloy has thermionic characteristics which are comparable with those of pure barium on tungsten, such an alloy cathode may have considerable practical interest.

3. PHYSICAL PROPERTIES OF OXIDE COATED CATHODES

A method for determining the true surface area of an oxide cathode was devised by Wooten and Brown in 1943. They pointed out that this area is deducible from the form of adsorption isotherms and measured the areas of a number of cathodes using ethylene at -183°C and -196°C and butane at -116°C. The results were self-consistent and had the form predicted by theory. The measured surface areas of supposedly identical cathodes, however, varied between fifteen and fifty times the apparent surface areas. Wooten and Brown concluded that this scattering was caused by inadequate control of the process of conversion of barium carbonate to oxide.

Measurements of the thermal emissivities of standard uncombined oxide cathode types were reported in 1941 by Moore and Allison. True temperatures were measured by Worthing's method and the results were checked against reflectometer measurements obtained by the method of Prescott and Morrison. Both the spectral and total emissivities were found to lie between 0.23 and 0.51 depending on the method of preparation of the cathode, and, in some cases, on the temperature.

The crystal structures of a variety of two component mixtures of alkaline earth oxides were investigated by x-ray and electron diffraction by Huber and Wagener in 1942. They verified earlier results which had demonstrated the formation of mixed crystals, and studied variations

in lattice parameter with component concentration. By electron diffraction they showed that barium oxide was absent in the outer few hundred atom layers of an active barium oxide-strontium oxide cathode.

4. ELECTRICAL AND THERMIONIC PROPERTIES OF OXIDE COATED CATHODES

An interesting study of the energy levels in oxide coated cathodes was published in 1940 and 1941 by Nishibori and his associates. By mounting a probe in the coating of an oxide cathode they were able to make simultaneous measurements of the conductivity and thermionic emission of the oxide. Since the conductivity depends exponentially on $eQ_1/(2kT)$ (cf. Fig. 1 or Fig. 6 of the author's 1939 review paper) and the thermionic emission depends exponentially on $e(Q_1/2 + \phi)/(kT)$, it was possible from these measurements to deduce the values of Q_1 and ϕ . Typical results obtained were as follows:

Before activation	After activation
$Q_1 = 1.68$ ev	$Q_1 = 1.40$ ev
$\phi = 0.27$ ev	$\phi = 0.28$ ev.

Since ϕ remained effectively constant, it is evident that activation of the cathode was caused by changes in the body of the oxide and not to any surface changes. Nishibori and his associates then proceeded to measure the photoelectric work function which should be given by $(Q_1 + \phi)$.² For an activated cathode they obtained a value of 1.66 ev in good agreement with the value of 1.68 ev deducible from the thermionic results.

Several other studies of oxide cathode work functions have appeared since 1940. Fan (1943) studied the variation of the A factor of a cathode with work function during activation over a range of work functions from 1.5 to 1.8 ev and obtained a more or less linear relation as did the classical authorities. Sproull (1945) measured work functions for impulse operation about 0.2 ev higher than the work functions observed for the same cathodes during steady operation. Sano (1941) described an ingenious method for meas-

uring the work function of a cathode under zero field conditions, by studying the plate current of a diode in the retarding field range. This method was checked for metallic cathodes and proposed for use on oxide cathodes, but no oxide cathode results were reported.

The Schottky effect for oxide cathodes was investigated by Fan (1943) and Sproull (1945). Schottky lines with slopes 3.7 and 2.5 times the theoretical slope were obtained by these two investigators, respectively. Presumably their cathode surfaces differed in roughness.

Early work on determination of oxide cathode temperatures from the shape of retarding field-emission current curves led to temperatures which were too high by as much as fifty per cent. The errors possible in such measurements were rather carefully investigated by Heinze and Hass (1938) who showed that temperatures of oxide cathodes can be determined by this method with errors of the order of one percent if the major sources of error are eliminated. Fan (1943) checked this conclusion by deductions of oxide cathode temperature which agreed within five percent with thermocouple measurements over the temperature range from 582°K to 951°K.

Generally qualitative electron microscope studies of the surface distribution of emission from oxide cathodes were reported by Fox and Bailey (1941) and Mecklenburg (1942). The surface distribution was found to be very non-uniform, an observation already well-authenticated by previous work on oxide cathodes. Mecklenburg noted that some of the emission centers were very small, in some cases of the order of 400 Angstrom units in diameter.

The emission of negative ions from oxide cathodes which plagues operators of magnetically controlled cathode-ray tubes received further attention from Bachman (1940) and Schaefer and Walcher (1943). Bachman explained some of the anomalous effects associated with magnetic resolution of negative ions in an oscilloscope tube as due to secondary effects at accelerating or focusing electrodes. Schaefer and Walcher performed a refined mass spectrometer analysis of the negative ions and reported a formidable list of ion masses. Unfortunately, their tube was not sealed off and some of the ions which they detected may not appear in

² Because of an unfortunate error in the author's 1939 review paper the photoelectric work function was said to be given by $(Q_1/2 + \phi)$. It is evident from Fig. 1 that the correct value is $(Q_1 + \phi)$.

conventional tubes. They made further observations relating negative ion emission, particularly of oxygen, to cathode activity.

Knowles and McNall (1941) described extensive experiments on cathode sparking when high currents were drawn from oxide cathodes in the presence of mercury vapor. They made careful measurements of tube drop before breakdown of time elapsed before breakdown occurred, and of breakdown current.

5. TIME EFFECTS

Johnson (1944) observed that bombardment of an oxide cathode by electrons from an external source produced changes in the electron emission from the cathode which decayed rather slowly after the bombarding current was turned off. The anomalous current was not attributable to temperature changes but may have its explanation in the electrolytic transport of barium which is considered to be responsible for the normal decay effect.

The normal decay effect was noted by Andrianov (1944) who was able to draw currents as high as 30 amperes per sq. cm for periods of a few microseconds from oxide coated cathodes. Fan (1943) was unable to detect any decay effect whatever. This interesting observation is, perhaps, to be associated with the fact that his cathode, from his account, appears to have been incompletely activated. A detailed study of the normal decay effect was published by Sproull in 1945 who outlined and expanded a theory of the effect. He showed that the effect was independent of cathode thickness and concluded that it depended, rather, on crystallite size. Also he discussed the "flicker" effect observed by J. B. Johnson and others and suggested that this effect might be associated with the decay effect since the time constants involved were of the same order of magnitude.

6. BIBLIOGRAPHY

The parentheses after each paper listed below indicate the feature of the paper of particular interest in connection with oxide coated cathodes.

1938.

J. Ewles, Proc. Roy. Soc. A167, 34 (1938) (Luminescence of alkaline earth oxides).

W. Heinze and W. Haas, Zeits. f. tech. Physik 19, 166 (1938) (Determination of oxide cathode temperature from retarding field emission characteristics).

1939.

J. P. Blewett, J. App. Phys. 10, 668 and 831 (1939) (Review article).

N. D. Morgulis, Tech. Phys. (USSR) 9, 853 (1939) (in Russian) (Theory of secondary emission from composite cathodes).

1940.

C. H. Bachman, J. App. Phys. 11, 83 (1940) (Formation and focusing of negative ions in cathode-ray tubes).

J. A. Becker and G. E. Moore, Phil. Mag. 29, 129 (1940) (Thermionic emission, migration, and evaporation of barium on tungsten).

K. Brünig, Physik. Zeits. 41, 285 (1940) (Thermionic properties of barium-copper alloy coatings on tungsten).

B. Gysae and S. Wagener, Zeits. f. Physik 115, 296 (1940) (Work functions of cathodes with non-uniform emitting surfaces).

E. Nishibori and H. Kawamura, Phys. Math. Soc. Japan Proc. 22, 378 (1940) (Conductivity and thermionic emission of oxide coated cathodes).

B. N. Srivastava, Proc. Roy. Soc. A175, 26 (1940) (Thermal ionization of barium vapor).

P. V. Timofeev and R. M. Aranovich, J. Tech. Phys. (USSR) 10, 32 (1940) (in Russian) (Oxygen-barium and oxygen-magnesium surfaces for secondary emitter applications).

Y. Uehara and M. Takahasi, Bull. Chem. Soc. Japan 15, 15 (1940) (Quantum mechanical theory of electron emission from oxide coated cathodes).

1941.

G. W. Fox and F. M. Bailey, Phys. Rev. 59, 174 (1941) (Electron microscope study of oxide coated cathode emission).

D. D. Knowles and J. W. McNall, J. App. Phys. 12, 149 (1941) (Sparking of oxide coated cathodes in mercury vapor).

G. E. Moore and H. W. Allison, J. App. Phys. 12, 431 (1941) (Spectral and total thermal emissivities of oxide coated cathodes).

N. D. Morgulis, Bull. Acad. Sci. (USSR) (Ser. Phys.) (in Russian) (Izvestia Akademia nauk SSSR, Seria fizicheskaya) 5, 536 (1941) (Nine-page review of oxide coated cathode fundamentals).

E. Nishibori, H. Kawamura, and K. Hirano, Phys. Math. Soc. Japan Proc. 23, 37 (1941) (Photoelectric work functions of oxide coated cathodes).

C. H. Prescott, Jr., "The pyrometry of oxide coated cathodes" in *Temperature, its Measurement and Control in Science and Industry* (Reinhold Publishing Corporation, New York, 1941), p. 1199.

S. Sano, Electrotech. J. (Japanese) 5, 78 (1941) (A new method for measuring work functions of oxide coated cathodes).

1942.

- K. H. Geyer, *Aan. d. Physik*, **42**, 241 (1942) (Secondary emission from alkaline earth oxides).
 H. Huber and S. Wagener, *Zeits. f. tech. Physik* **23**, 1 (1942) (X-ray and electron diffraction studies of oxide coated cathodes).
 W. Mecklenburg, *Zeits. f. Physik* **120**, 21 (1942) (Electron microscope study of oxide coated cathode emission).

1943.

- H. Y. Fan, *J. App. Phys.* **14**, 552 (1943) (Thermionic emission from oxide coated cathodes).
 G. Herrmann and S. Wagener, *Die Oxydkathode*, Physikalische Grundlagen (J. A. Barth, Leipzig, 1943), Vol. I (124 pages) (This book is not yet available in the United States).
 H. Schaefer and W. Walcher, *Zeits. f. Physik* **121**, 679 (1943) (Mass spectrometer analysis of negative ions emitted from oxide cathodes).

- C. F. Veenemans, *Ned. Tijd. Naturkunde*, **10**, 1 (1943) (in Flemish). (Sixteen-page review on oxide coated cathodes).
 L. A. Wooten and C. Brown, *J. Am. Chem. Soc.* **65**, 113 (1943) (Surface area of oxide coated cathodes by adsorption of gas at low pressures).

1944.

- M. Andrianov, *Bull. Acad. Sci. (USSR) (Ser. Phys.)* (in Russian) (*Izvestia Akademii nauk SSSR, Seria fizicheskaya*) **8**, 290 (1944) (One-page discussion of impulse emission from oxide coated cathodes).
 J. B. Johnson, *Phys. Rev.* **66**, 352 (1944) (Anomalous emission from oxide coated cathode bombarded by electrons).

1945.

- R. L. Sproull, *Phys. Rev.* **67**, 166 (1945) (Short time emission from oxide coated cathodes).

The Pulsed Properties of Oxide Cathodes*

EDWARD A. COOMES†

Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

A survey of the results obtained in an experimental study of the pulsed properties of oxide cathodes is presented. Pulsed measurements reveal that unusually large electron currents are available in microsecond pulses, and that several other phenomena are modifications of the d.c. properties, namely (a) sparking and (b) pulse temperature rise. Sparking may be either current limited or voltage limited, depending upon cathode materials and life. The pulse temperature rise also depends upon materials and life and is indicative of the nature of the cathode resistance. Evidence for a layer structure of the oxide cathode can be drawn from pulsed data.

I. INTRODUCTION

THE fact that oxide cathodes have interesting time-dependent properties was clear before the war. J. Blewett pointed this out in his 1939 paper¹ on the time changes in oxide cathodes. O. Schade, in his discussion of rectifier analysis,² stated that transient peak currents of 25 amp./cm² had been observed from well-activated oxide cathodes. A good phenomenological picture of cathode sparking was also given, with a clear distinction between the two types.

The extensive use of pulsed electron tubes during the war invoked interest in the properties

of oxide cathodes under pulsed conditions, since it appeared that unusually large electron currents were available in short pulses. Before 1942 at Radiation Laboratory, J. McNall³ had obtained microsecond pulses of emission of more than 1 amp./cm² at 850 deg. C, under space charge limited conditions. He used a diode having an internal anode, and his measurements were taken at a relatively high pulse recurrence frequency (p.r.f.), so that currents were limited by anode dissipation.

Ramsay at Bartol, using single pulses, observed two types of decay in the pulsed emission from oxide cathodes, the first type having a decay time of about 10⁻⁶ second, and the second, a decay time of the order of milliseconds. The short time decay, which is of interest for microsecond pulse phenomena, was observed in aged cathodes

* This paper is based on work done for the Office of Scientific Research and Development under contract OEMar-262 with the Massachusetts Institute of Technology.

† Now at Department of Physics, University of Notre Dame, Notre Dame, Indiana.

¹ J. P. Blewett, *Phys. Rev.* **55**, 713 (1939).

² O. Schade, *Proc. I.R.E.*, **31**, 341 (1943).

³ J. W. McNall, M.I.T. Doctor's Thesis (1942).

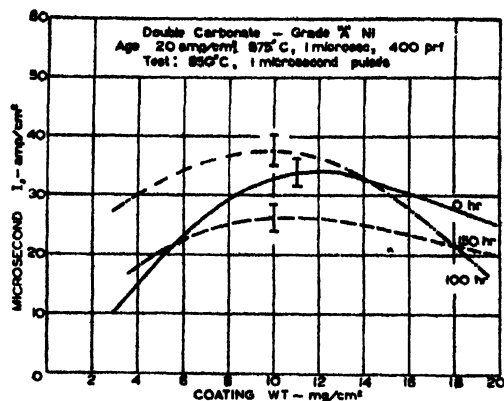


FIG. 1. The dependence of microsecond pulsed emission on coating weight for plain oxide cathodes. The maximum spread of experimental points is indicated.

but could not be found in very active cathodes. R. Sproull,⁴ using pulse widths in the range 0.2 to 300 microseconds, observed experimentally that the rate of decay of emission was independent of cathode coating thickness.

In the experiments that followed the work of McNall at Radiation Laboratory, microsecond pulses of emission space-charge-limited⁵ up to about 5 amp./cm² at 850°C were obtained from oxide cathodes of the prewar commercial variety, using an early design of test diode with an external water cooled anode. About this time the screen cathode⁶ emerged from the technological developments of oxide cathodes for pulsed transmitter tubes, and with this type of structure microsecond emissions of at least 35 amp./cm² at 850°C were consistently obtained in test diodes. Apparently there was nothing very fundamental in the screen structure that could account for the difference in emission between the two types of cathodes, and in making a comparison only two significant differences were noted at the time: (a) the coating weight per unit area was much larger on screen cathodes, having a value of about 20 mg/cm² compared with about 6 mg/cm² for the plain type and (b) because of the unique preparation technique for screen cathodes, the density of coating was about twice that ordinarily obtained with plain coatings. Tests were run on a

⁴ R. Sproull, Phys. Rev. 67, 166 (1945).

⁵ The maximum value of space-charge-limited emission is taken at the point where the cathode emission characteristic plotted on $\frac{1}{2}$ -power paper deviates from the computed Langmuir-Child line.

⁶ The screen cathode is made by firmly bonding nickel mesh to a supporting nickel sleeve and impregnating the interstices of the mesh with cathode coating.

series of grade "A" nickel plain sleeves coated with a double-carbonate suspension, and a good dependence of pulsed emission on weight was obtained, with the highest value of microsecond pulsed emission at about 10–12 mg/cm². The results obtained by J. Buck in his studies during the life of a series of plain cathodes is given in Fig. 1. This shows that the dependence of emission on coating weight is maintained for at least 150 hours of operation at 850°C. The result is also significant in that it was not found to apply in the case of d.c. emissions; no systematic dependence of d.c. emission on coating weight was observed.

From the time that pulsed oxide cathodes became technologically important there was hope that correlations could be made between the d.c. and pulsed properties. It appeared from early diode experiments that definite correlations of a simple nature did not exist. About all that could be inferred from correlation experiments was that usually good pulsed cathodes were also good d.c. cathodes, but the reverse need not be true. For example, in a certain life test three groups of cathodes (a), (b), and (c) with selected structural differences all gave $\frac{1}{2}$ amp./cm² d.c. emission at 800°C for over 500 hours of life. Group (a) gave a microsecond pulsed emission of 60 amp./cm², group (b) 20 amp./cm², and group (c) $\frac{1}{2}$ amp./cm², for the same life test conditions.

II. PARAMETERS FOR DESCRIBING THE PULSED PROPERTIES OF OXIDE CATHODES

Tests designed to give extensive information on the pulsed properties of oxide cathodes reveal phenomena and limitations which are usually not found by d.c. methods. Two effects are particularly serious in tending to limit and confuse the interpretation of pulse data.

(a) The Temperature Rise

When large currents are drawn in pulses with a relatively high pulse recurrence frequency from an oxide cathode, there is a large rise in the average temperature of the cathode, as illustrated in Fig. 2. This effect may be minimized by going to lower duty cycles; with microsecond pulses at 60 p.r.f. the effect is not observable by ordinary methods of optical pyrometry on usual oxide cathodes.

(b) Sparking

The second, and probably most serious, limitation encountered in the pulsed observations on oxide cathodes is the phenomenon of sparking. In this case the electron current from the cathode is limited by the physical transfer of particles of oxide coating from the cathode sleeve to the anode, accompanied by a large increase in electron current with an extremely steep wave front. The sparking value of current on pulsing usually occurs before temperature saturation of emission is reached, and for the most active cathodes it will occur before the cathode emission deviates from space charge limitation. An example of this case is given in Fig. 3. With this situation the problem of improving the pulsed thermionic emission immediately transforms to the problem of improving sparking properties. Experiment indicates that the space charge limited value of pulsed sparking current is higher for shorter pulse widths; for example, during the normal life of an active oxide cathode the sparking current for 10 microsecond pulses is about $\frac{1}{3}$ the value obtained for 1-microsecond pulses, at a low pulse recurrence frequency. This means that to measure the highest instantaneous value of electron current available from an oxide cathode well-defined voltage pulses below a microsecond in width must be used.

While sparking and temperature rise place limitations on the pulsing of oxide cathodes, the phenomena related with these two effects have good reproducibility and may be used to define parameters of measurements for different cathode types.

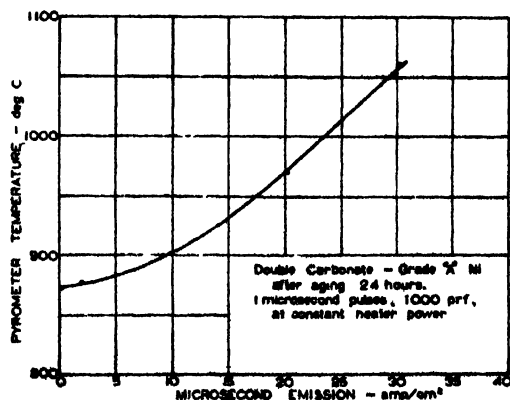


FIG. 2. The rise in temperature with increase of magnitude of pulse current for an oxide cathode.

FIG. 3. Space charge limited sparking

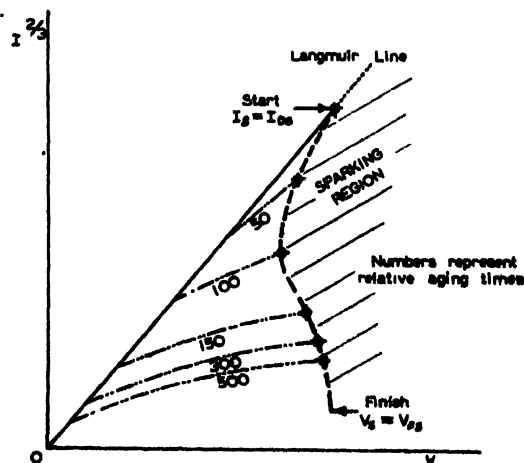
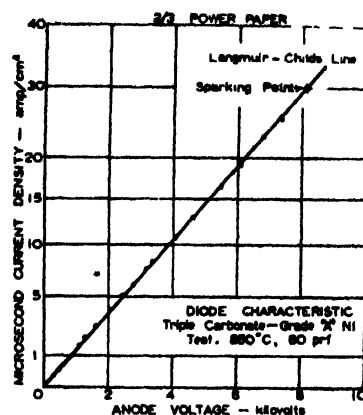


FIG. 4. Sparking diagram for a pulsed oxide cathode.

1. The Sparking Points

These are defined by the $I-V$ diode values at which the electron current from the cathode is limited by sparking. Data of this type cannot be taken on d.c., but with short pulses at a low recurrence rate not only can the data be reproducibly obtained but important connections with the chemical and physical details of the cathode structure seem to exist. A comprehensive representation of the sparking phenomena may be arrived at by drawing I^2 versus V curves at various stages in the life of the cathode as illustrated in Fig. 4. This diagram represents the probable life cycle of a well-aged, well-activated oxide cathode for 1-microsecond pulsed emission at a low recurrence rate. While it has been recognized before² that there are two types of sparking, this is brought out clearly in the study of the pulsed sparking currents. Differentiation

between the two types may be made in the manner stated below.

(a) *Space-charge limited sparking* occurs early in the life of the cathode when the current is limited to a value $I_s = I_0$ on the space charge line (see Fig. 4). As a result of experiment, an I_0 value of at least 50 amp./cm² at 800°C and 1 microsecond may be assigned to cathodes prepared with a good double carbonate coating on grade "A" nickel, while values as high as 130 amp./cm² have been obtained under pulsed conditions with purer nickels as base metals. The space-charge limited sparking current is an important parameter for pulsed oxide cathodes because it is the highest value of electron current that can be taken from the cathode, and because it seems to bear a definite relation to the material system which comprises the cathode.

(b) *Voltage limited sparking* takes place late in the life of an oxide cathode when the diode sparking occurs at a value of anode voltage $V_s = V_{fs}$ (see Fig. 4). This may mean that sparking takes place at a critical value of field at the cathode surface, and, therefore, it is an important parameter since it may depend in a fundamental way upon the nature of the oxide at the cathode surface. In experiments on cathodes prepared with strontium carbonate or double carbonates, the computed field corresponding to V_{fs} was found to be 40–60 kv/cm, while for cathodes prepared with pure barium carbonate the field was in the range 100–125 kv/cm. Using cathode coating prepared with alternate layers of the single carbonates, results for sparking field checked the above ranges, giving the low value of sparking field when strontium oxide was formed on the surface, and the higher value when barium oxide was on the surface.

While sparking of the two types may usually be distinguished for oxide cathodes during a life test, there is experimental evidence to indicate that in some cases sparking may be dependent on the state of the anode, which would prevent the observation of either one or both of the two fundamental types already described.

2. The Maximum Space-Charge-Limited Emission

If the cathode does not spark on the space charge line, then the point I_0 at which the plotted experimental data first deviate from the theo-

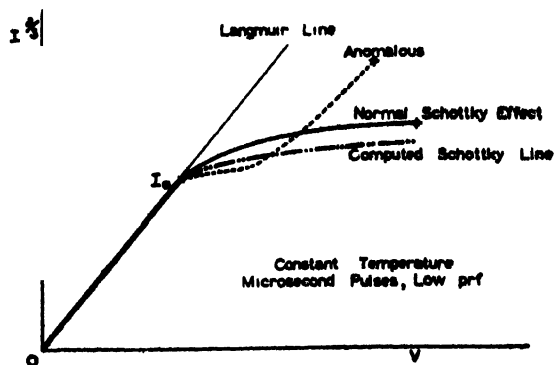


FIG. 5. Typical emission characteristics of an oxide cathode pulsed at a constant temperature.

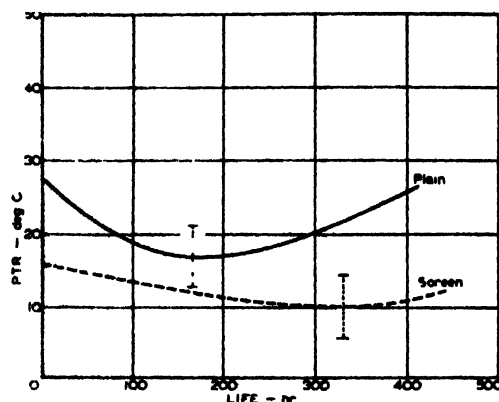


FIG. 6. Pulse temperature rise as a function of life.

retical Langmuir-Child line is the maximum electron current obtained from the cathode under space charge limited conditions. If this point is interpreted as the maximum thermionic emission current that can be obtained at a constant temperature with zero field at the cathode surface then this is an important parameter since it corresponds to the value of current given by Richardson's equation.⁷ Figure 5 illustrates the manner in which I_0 is obtained. This figure also illustrates the general trends of pulsed data above the space charge limited value. At a given time in life, there will be observed one of two types of behavior: cathodes will either have $I-V$ characteristics that follow the normal Schottky equation but possibly with a temperature different than the observed value, or there will be an anomalous behavior which is much more pro-

⁷ A series of values of I_0 at various temperatures for average good cathodes prepared from double-carbonate coating on grade "A" nickel, gave a computed value of about 1 volt for the work function.

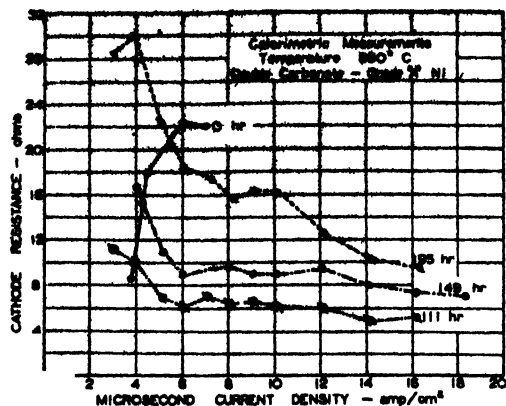


FIG. 7. Oxide cathode resistance measured under pulsed conditions. (Data taken by J. Buck.)

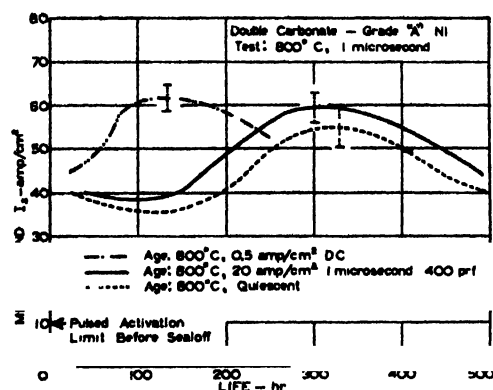


FIG. 8. The effect of duty cycle on the life of a pulsed oxide cathode.

nounced than in the case of the well-known d.c. effect.⁸

3. The Pulse Temperature Rise

The average temperature rise which occurs on pulsing at higher recurrence rates may be used as a parameter, when measured in a standard way. This pulse temperature rise (ptr) may be specified at a given current level and duty cycle. For example, when an average oxide cathode is pulsed at 10 amp./cm², 1-microsecond pulses, 1000 p.r.f., the pyrometer temperature may show an increase of from about 5 to 60°C above the initial quiescent temperature at zero current. This ptr may be studied over the life of the cathode, or it may be compared for different cathode systems. Experiment indicates that, in general, during the life of a cathode the ptr falls

⁸ See A. L. Reimann, *Thermionic Emission* (John Wiley and Sons, New York, 1934), Chapter VI.

to a minimum and then increases for later times. Figure 6 shows that the ptr for an average good screen cathode is always lower than the ptr for an average good plain oxide cathode. If the extra cathode heat which causes the ptr is attributed to the RI^2 loss resulting from the passage of current through the resistance of the cathode, then the ptr is an indication of the cathode resistance, and the resistance R may be determined experimentally by a calorimetric method, based on the equation

$$W_H + RI^2bf - \phi Ibf = \epsilon \sigma T^4 + \Sigma,$$

where

- W_H = cathode heater power input,
- I = pulsed current,
- bf = duty cycle, which is pulse width times p.r.f.,
- ϕ = thermionic work function,
- ϵ = total emissivity,
- σ = Stefan's constant, and
- Σ = conduction and convection heat losses.

In this method at a chosen current I , and for two different values of duty cycle, values of W_H are alternately set to give the same temperature. Substituting these data in the equation (neglecting ϕI and Σ) gives two simultaneous equations from which a value of R corresponding to a given I may be computed. A typical plot of resistance *versus* current obtained in this manner during the life of a cathode is given in Fig. 7. Since there is a striking resemblance between the R *versus* I curve for an oxide cathode and this same characteristic curve for a blocking layer rectifier in the high resistance direction, the hypothesis first suggested by Reimann,⁹ that a high resistance contact between coating and the base metal might be responsible for the observed variations of cathode resistance with current, deserves consideration. Measurements were made using probes embedded in the coating at various distances from the base metal,¹⁰ and it was found that at least in the case of cathodes prepared with double carbonates on pure and Cr plated nickel, the resistance by probe measurements had the same type of current dependence as found by calorimetric measurements, and that the resistance at zero life at the metal-to-coating interface

⁹ A. L. Reimann and R. Murgoci, *Phil. Mag.* 9, 440 (1930).

¹⁰ A. Fineman and A. Eisenstein, "Studies of the interface of oxide cathodes," *J. App. Phys.* 17, 663 (1946).

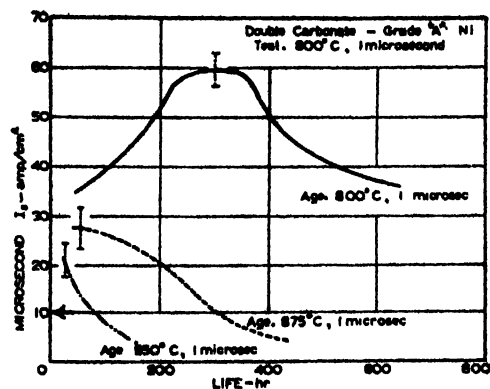


FIG. 9. The effect of operating temperature on the life of a pulsed oxide cathode.

at operating temperature of 850°C was in some cases an order of magnitude greater than the resistance of the coating.

III. LIFE PROPERTIES AND THE EFFECTS OF MATERIALS

Interesting phenomena in the pulsed studies of oxide cathodes come out of life testing, and it becomes evident that before any comparison of the pulsed properties of various oxide cathodes can be made one must realize just how these properties vary during the life of a single cathode. As an example, some life studies of the cathode sparking current are given. Figure 8 is an average plot of the microsecond sparking current I_s as a function of life on cathodes prepared with double carbonate coatings on grade "A" nickel, each cathode having a weight of 12 mg/cm² and having undergone a good standard processing schedule. While this diagram shows particularly the effect of duty cycle, the nature of the curves is general for life test data taken under pulsed conditions. Each curve has an initial rise, a relatively flat portion, and then a definite falling off with life, the trends defining three different regions of activity in the life of the cathode.

Region I: The cathode is still apparently undergoing aging so that the activity is not constant. In the experiment from which these data were taken the activation was purposely stopped at 10 amp./cm² to allow the observation of this effect.

Region II: This might be defined as the normal life of the cathode, since in this region the activity is at its highest measurable value and remains practically constant throughout the period.

Usually, the sparking current density over this period of life remains near its space charge limited value I_{0s} .

Region III: This is the region of constant sparking voltage, since over this period of life sparking takes place at practically the same value of voltage, V_s equal to V_{fs} . As the cathode activity falls with life, the sparking current density will show a decline, since it is measured at a constant anode voltage.

The effect of operating temperature on the sparking properties of a pulsed oxide cathode is shown in Fig. 9. When compared with Fig. 8 this illustrates particularly that under pulsed conditions operating temperature exerts a much stronger influence on the obtainable level of cathode activity than does the duty cycle.

Life tests have consistently indicated that the effect of cathode base metal on the pulsed properties of oxide cathodes is a very strong one, and to date the nickels of higher purity than grade "A" have given higher values of pulsed emission. Figure 10 gives a comparison of results with a good electrolytic nickel¹¹ base metal against the results obtained with a good grade "A" nickel.

IV. THE LAYER STRUCTURE OF OXIDE CATHODES

Several types of evidence come out of pulsed measurements that might indicate a layer structure for an oxide cathode as suggested by Fig. 11. There are the probe measurements of cathode resistance which show a comparatively high resistance in the interface region. The variation of

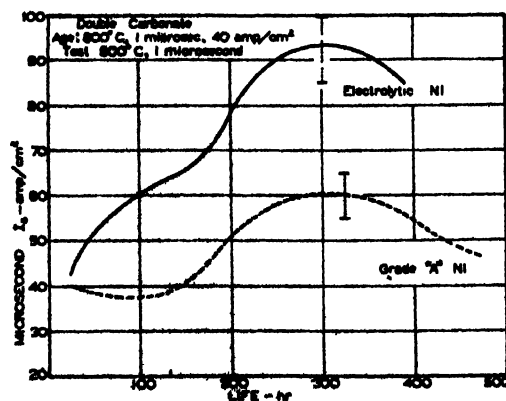


FIG. 10. A comparison of pulse emissions with pure and grade "A" nickel cathode base metals.

¹¹ The electrolytic nickel used in these tests was supplied by Mr. E. M. Wise of International Nickel Company.

this resistance with current, and the temperature dependence of this relation are those that would be expected¹² if the oxide cathode had a structure similar to a blocking layer rectifier. This model also lends itself to a straightforward explanation of high pulsed emissions when microsecond voltages pulses are applied. The conduction band of the semiconducting oxide, which is filled by the thermally excited electrons from the impurity levels, might act as a reservoir of electrons from which large transient currents could be taken at the oxide-vacuum surface. The emission limitations of the high resistance blocking layer at the interface may be overcome by the production of high fields in this blocking layer which in turn cause a copious emission of electrons at the base metal surface by means of the Schottky effect or even by field emission. Although theoretically a blocking layer that could withstand such a field can be formed by simply placing a semiconductor in contact with a metal, it has been observed that metal-oxide interfaces both in the case of copper oxide rectifiers¹³ and in oxide cathodes are usually compounds that are different than the main body of the semiconductor. For example, A. Eisenstein,¹⁰ by the use of x-rays has found complicated barium compounds at the interfaces of cathodes made on nickels containing small amounts of selected metallic impurities.

There are other experimental facts that indicate the cathode interface as a possible important factor in any consideration of electron emission mechanisms from oxide cathodes. Extensive life tests have indicated that for pulsed emission the selection and treatment of the base metal before coating and the details of the subsequent breakdown schedule are more important than the choice of carbonate coating. During processing, not only can impurities initially contained in the base metal reduce the oxide coating to produce free alkaline earth metal, but it is possible that they also effect the rate, type, and extent of the compound formation at the interface. For example, it was observed that ordinarily very thin coatings never produce good pulsed emissions; however in an experiment performed by A. Fineman, coatings of the order of 1 mg/cm² were converted in an atmosphere of several microns

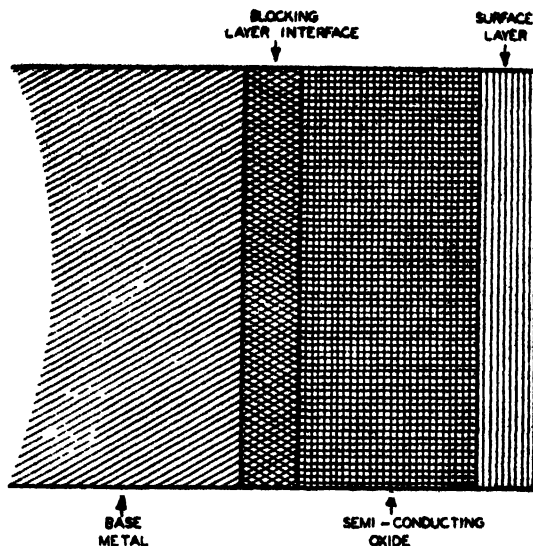


FIG. 11. A layer model for oxide cathodes.

pressure of the breakdown gases from a larger cathode heated simultaneously, and pulsed emissions of 25 amp./cm² were obtained. Details such as sharpness and completeness of conversion entering as an essential requirement for obtaining high pulsed currents, may point to a critical pressure-temperature value at the base metal surface during conversion as being important in the cathode forming process. There is also the direct experimental evidence obtained by J. Buck, that preoxidation of certain nickel base metals produces cathodes that give higher pulsed emissions than the same base metal in an unoxidized condition.

Rather good evidence has been obtained for the existence of the surface layer of an oxide cathode. While experiments that have been carried out to give direct evidence for a monolayer of barium have not been conclusive,^{14,15} these same experiments did reveal that during the life of a barium-strontium oxide cathode there is formed on the cathode surface a thin layer of pure strontium oxide. Life test experiments in conjunction with x-ray measurements by A. Eisenstein¹⁶ indicate that this layer reaches a thickness of about 10⁻⁵ cm roughly at the time the sparking current becomes voltage limited, or at the beginning of Region III of the life diagram. It has been noted

¹² N. F. Mott, *Proc. Roy. Soc. A* 171, 27 (1939).

¹³ L. O. Grondahl, *Rev. Mod. Phys.* 5 (1933).

¹⁴ J. A. Darbyshire, *Proc. Phys. Soc.* 50, 635 (1938).

¹⁵ H. Huber and S. Wagener, *Zeits. f. tech. Physik* 23, 1 (1942).

¹⁶ A. S. Eisenstein, *J. App. Phys.* 17, 654 (1946).

that this is also the same time at which the ptr reaches its minimum value during life. It is thus apparent that the available pulsed emission can be limited either by conditions at the surface or at the interface, and the nature of this limitation can only be ascertained by a careful consideration of the effect of the operating temperature and the thermochemistry of the cathode structure as a whole at this temperature.

The author wishes to thank the members of the Cathode Research Group of Radiation Laboratory for their cooperation in this wartime research. James Buck, Abe Fineman, and Albert Eisenstein were the staff members on whom fell the burden of the experimental program, and Earnest Parsons managed the test program that produced the enormous amount of data needed to complete the work.

A Study of Oxide Cathodes by X-Ray Diffraction Methods*

Part II. Oxide Coating Composition

A. EISENSTEIN†

Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

X-ray diffraction methods were used to investigate time changes which occur in the composition of oxide cathode coatings, initially equal molar (BaSr)O. Changes which occur in the composition of the bulk of the coating were detected by means of lattice constant measurements. A new method of analysis was developed to permit a determination of the variation of composition with depth below the surface. The bulk loss of BaO is primarily a function of the base metal used and the surface loss from the oxide is effected to a lesser extent. Possible correlations with thermionic emission are discussed.

PART I of this paper described in detail the x-ray diffraction method as applied to the study of oxide coated cathodes. This technique is particularly suited to the analysis of time changes in composition which occur in oxide cathode coatings. Because of the non-destructive nature of this method extensive studies can be made of single cathodes. As a tool, x-ray diffraction measurements reveal only the crystal structure and the crystallographic lattice constants of the sample, hence some *a priori* knowledge is required of the relationship between crystal structure and molar composition. The oxides BaO and SrO, which have the same crystal structure, are reported^{1,2} to form true solid solutions, the lattice constant of which varies continuously and linearly with composition.

By means of x-ray diffraction analyses, Benja-

min and Rooksby³ found that the compositions of several different (BaSr)O oxide solid solutions changed by a few percent during the thermal activation of the cathode because of the preferential evaporation loss of the BaO component. They also report a considerable bulk loss of BaO at the "end of life" of a group oxide coated cathodes.

Gaertner,⁴ Darbyshire,⁵ and Huber and Wagener,² using electron diffraction analyses, agree that in activated cathodes the bulk composition of the oxide solid solution, is not homogeneous throughout but that the surface usually consists of pure SrO. Burgers¹ has shown that a cathode "flashed" at 1225°C will develop a relatively thick BaO deficient surface layer in 30 minutes, this layer being detected by an asymmetry in the shape of the diffracted x-ray lines.

In this paper, variations in coating composition which are detected by changes in the crystal lattice constant are presumed to be a character-

* This paper is based on work done for the Office of Scientific Research and Development under Contract OEMar-262 with the Massachusetts Institute of Technology.

† Now at the Research Laboratory of Electronics, Massachusetts Institute of Technology.

¹ W. G. Burgers, *Zeits. f. Physik* 80, 352 (1933).

² H. Huber and S. Wagener, *Zeits. f. tech. Physik* 23, 1 (1942).

³ M. Benjamin and H. P. Rooksby, *Phil. Mag.* 15, 810 (1933).

⁴ H. Gaertner, *Phil. Mag.* 19, 82 (1935).

⁵ J. A. Darbyshire, *Proc. Phys. Soc.* 50, 635 (1938).

istic of the bulk of the coating and are termed bulk changes. Surface changes refer to those variations in composition which occur near the external surface of the coating and which are detected by asymmetries in the shape of the diffraction lines.

EXPERIMENTAL

The experimental techniques used to obtain x-ray diffraction patterns of the chemically unstable oxides of barium and strontium have been described in detail.⁶ X-ray diffraction patterns were taken of the oxide cathode coatings in a 4.70-cm radius camera using filtered $\text{CuK}\alpha$ - or $\text{CoK}\alpha$ -radiation. The following groups of cathodes were studied.

Group 1

In a preliminary survey, a group of 16 cathodes which had reached the "end of life" were examined to determine, if possible, what physical changes in the oxide coating could be correlated with the decline in thermionic activity. This group consisted of standard⁷ test cathodes run in the laboratory diode.⁸ For the 11 cathodes which were operated and tested⁹ under pulsed conditions of 1 $\mu\text{sec.}$, 400 and 50 p.r.f., respectively, the "end of life" was defined as that time at which the sparking current⁷ fell below 10 amp./cm². For the 5 cathodes operated under d.c. conditions the "end of life" state was reached when the emission current fell below 0.1 amp./cm² at an anode voltage capable of giving initially, 1 amp./cm² space charge limited emission. Life tests were carried out at constant, uncorrected, pyrometered temperatures in the range 800°C to 950°C at which the initial pulsed cathode emission currents at sparking were in excess of 50 amp./cm² and the d.c. cathode emission levels were 1 amp./cm².

⁶ A. Eisenstein, "A study of oxide cathodes by x-ray diffraction methods," *J. App. Phys.* 17, 434 (1946).

⁷ The cathodes were coated approximately 10 mg/cm², with an equal molar solid solution carbonate (BaSr)CO₃, (Rathcon C51-2); 8-mm coated length on 3-mm $\times$ 14-mm indirectly heated sleeves. Pure nickel, grade A nickel, and nickels containing added amounts of chromium and silicon were used as base metals.

⁸ A 9.5-mm external Kovar anode tube. Anode outgassed by electron bombardment in a pre-processing treatment prior to mounting the test oxide cathode. Anode water cooled during test and life run periods.

⁹ E. A. Coomes, "The pulsed properties of oxide cathodes," *J. App. Phys.* 17, 647 (1946).

Group 2

To investigate the effect of temperature, time, and base metal, on coating composition, twelve monodes containing standard cathodes were processed and life tested. Eight of these had as a base metal, electronic grade A nickel and four a pure nickel, RCA N-81. These monodes were evacuated using a three-stage oil diffusion pump and baked for one hour at 450°C. The cathodes were processed and the tubes sealed off at a pressure of less than 10^{-7} mm. These cathodes were operated quiescent, i.e., glowed without emission current at either 800°C or 875°C for periods of from 50 hours to 1200 hours. At the end of this time interval the tubes were opened and x-ray diffraction patterns were taken of the cathode coating.

Group 3

Eight standard cathodes, prepared on the pure nickel, N-81, were processed in the laboratory diode. Five of these were operated for 40 hours in a quiescent state at a temperature in excess of 950°C and three were operated at a d.c. emission current of 0.5 amp. cm² at the same temperature and for the same time interval. X-ray patterns were then taken of these cathodes to determine the effect of a thermionic emission current on the change of oxide composition with time.

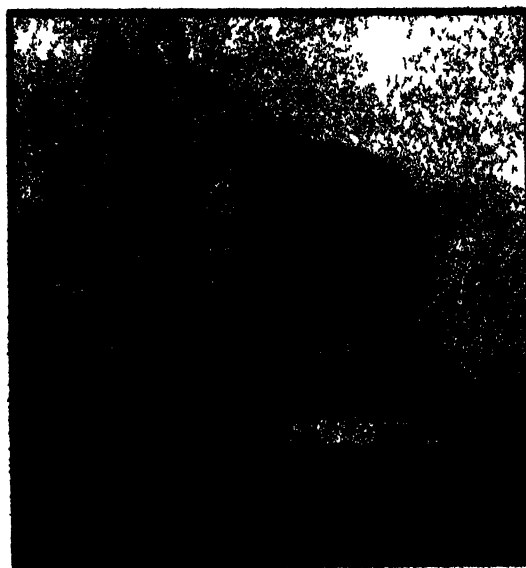


FIG. 1. Beryllium windowed, continuously pumped diode mounted at center of an x-ray camera.

Group 4

Three standard cathodes were operated in continuously pumped, beryllium windowed diodes.⁶ Figure 1 shows a diode of this type

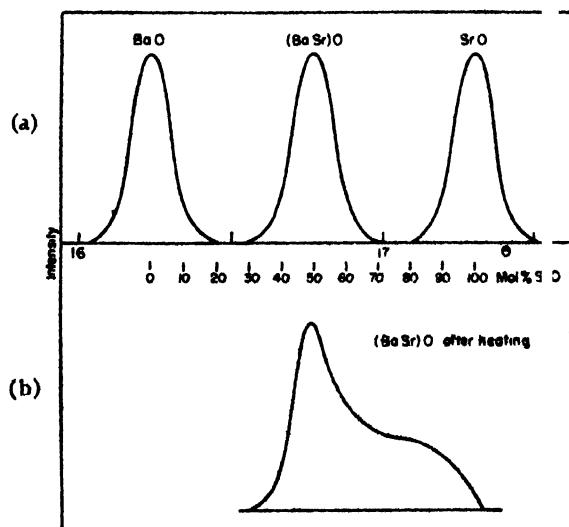


FIG. 2. Idealized 200 diffraction lines from (a) homogeneous phases BaO, (BaSr)O, and SrO and (b) an inhomogeneous solid solution, see text.

mounted at the center of the x-ray camera which is rigidly mounted with respect to the x-ray source. A pressure of less than 5×10^{-7} mm could be maintained in the vacuum system throughout the life test period. Diffraction patterns were taken at time intervals during the quiescent life test periods of these cathodes which in one case extended for 1200 hours. Analyses of the series of diffraction patterns thus obtained permitted evaluations of (1) bulk changes in the oxide coating composition and (2) surface changes in composition.

To check the results obtained in the continuously pumped diodes, one standard cathode was life tested in a sealed-off, beryllium windowed diode.⁶ The copper anode block of this tube was outgassed by electron bombardment in a pre-processing treatment. Following a baking period of one hour at 450°C the cathode was converted and activated, and the diode sealed off at a pressure of less than 10^{-7} mm. The diode was mounted in the x-ray camera and life tested at a quiescent temperature of 875°C .

METHODS OF ANALYSIS

In the absence of a detectable surface layer, the oxide solid solution was always homogenous throughout and produced relatively sharp diffraction lines. Figure 2a shows an idealized representation of the 200 diffraction lines of (1) BaO, (2) equal molar (BaSr)O, and (3) SrO. The intensities are plotted both as a function of half the scattering angle 2θ and the molar composition. The breadth of the curves has been adjusted to correspond to an average value observed experimentally. If an oxide coating, initially equal molar (BaSr)O is heated at a high temperature, a preferential loss of BaO will occur near the surface and the shape of the intensity curve will change from that of (BaSr)O of Fig. 2a to the asymmetrical form as indicated in Fig. 2b. The peak of the curve represents the bulk composition of the coating while the asymmetry corresponds to scattering from BaO deficient coating layers near the surface.

The lattice constant, determined from the peak positions, was used to evaluate the bulk molar composition of the oxide solid solution which was assumed to vary linearly² between $5.523\text{\AA}$ for BaO and $5.143\text{\AA}$ for SrO. An experimentally determined correction was required⁶ to compensate for line displacements due to absorption and to the large sample diameter. In spite of this correction it was occasionally observed that measurements of the lattice constant made from the diffraction lines of 2θ near 90° were not in agreement with measurements from the smaller 2θ lines, 111, 200, and 220. This discrepancy was believed to be caused by a variation of bulk composition with depth since in these patterns the large 2θ lines always gave slightly lower compositions of SrO than did the smaller 2θ lines. Only the 111, 200, and 220 diffraction lines were used in these analyses in order that the bulk composition always be determined in the same manner. The accuracy with which the bulk molar composition could thus be determined varied with the crystal size of the oxide and with its resulting effect on the "graininess" of the diffraction lines but was in general between two and five percent.

A method of analysis has been devised which allows a calculation of the composition vs. depth

function to be made from the experimental intensity vs. composition curves. It is assumed that the variation of composition with depth below the surface is a monotonic function between the limits 100 percent SrO at zero depth and a known concentration (50 percent SrO in Fig. 2b) at the maximum depth from which x-ray scattering is observed.

For purposes of analysis, the oxide is assumed to possess a layered structure, in any layer of which the oxide is homogeneous in composition. Figure 3 shows a structure of this type consisting of layers *A*, *B*, *C*, —, —, having respective thicknesses l_A , l_B , l_C , —. An x-ray beam of total power P_0 incident upon the structure gives rise to sets of diffracted beams of power P_A , P_B , P_C , —, —. Each set corresponds to one of the allowed *hkl* diffraction lines of the oxide solid solution. Since the composition and hence the lattice constant of each layer was assumed to have a unique value the diffracted beams P_A , P_B , P_C , —, —, will each be scattered at slightly different angles with respect to the incident beam, P_0 . The superposition of scattering from all layers gives rise to the asymmetrical diffraction line, Fig. 2b. When the composition increments distinguishing adjacent layers are sufficiently small, the experimentally observed diffraction lines can

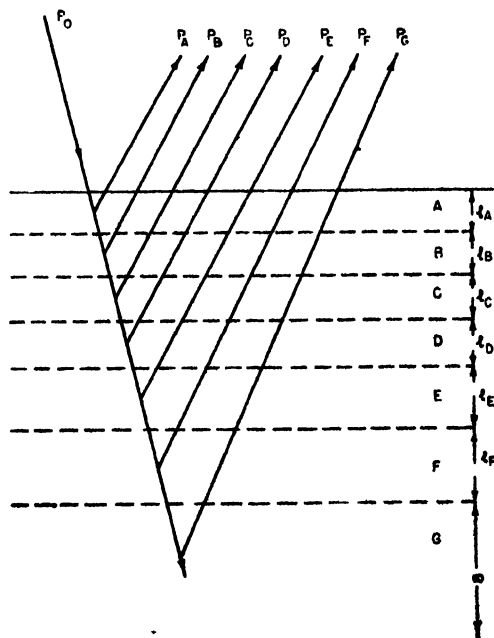


FIG. 3. Assumed layered structure of oxide coating at the external surface (see text).

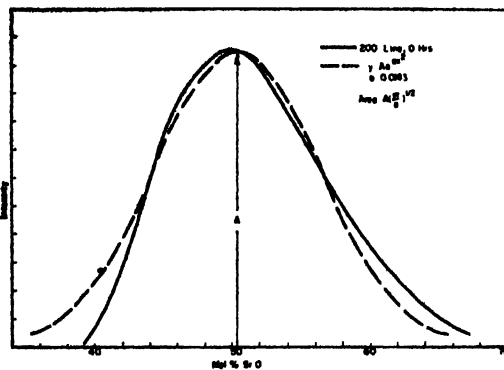


FIG. 4. Error curve fitted to experimental 200 diffraction line.

be synthesized by the summation of increments of power scattered by the various layers. Composition increments of five percent were found adequate to give a reasonable agreement with the experimental curves. These curves were synthesized by two methods which differed only in the form chosen to describe the angular distribution of intensity of the diffracted power increments. In one, the increments of diffracted power P_A , P_B , P_C , —, —, were represented by a set of Gaussian error curves, each having the same coefficient in the exponent as the original diffraction line from the completely homogeneous oxide (see Fig. 4). Error curves were of the form,

$$y = A \exp(-ax^2), \quad (1)$$

where *A*, the amplitude factor, is the height of the curve and *a* is the constant coefficient determined from the original diffraction line. Figure 4 shows an error curve fitted to the 200 diffraction line obtained from a homogeneous equal molar solid solution oxide. The obvious asymmetry of the diffraction line is caused^{10,11} primarily by absorption in the sample. The coefficient *a* is a function of oxide crystal size, beam divergence, and sample dimensions. When once determined for the cathode at zero hours life, this factor should be applicable to diffraction lines from the same cathode taken later in life but under identical experimental conditions, assuming that crystal growth does not occur during the life test period.⁶ With a knowledge of the coefficient *a*, the amplitudes *A_i* of the diffracted power

¹⁰ A. Taylor and H. Sinclair, Proc. Phys. Soc. **57**, 108 (1945).

¹¹ B. E. Warren, J. App. Phys. **16**, 614 (1945).

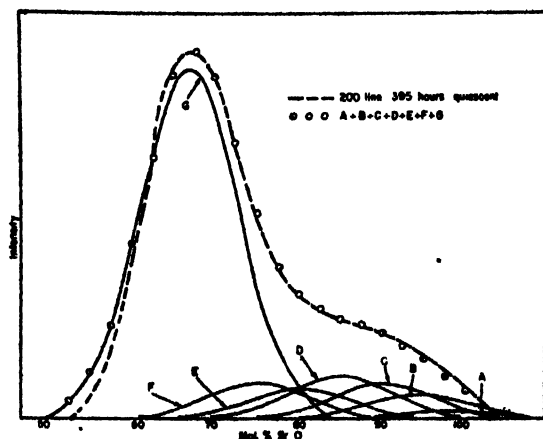


FIG. 5. 200 diffraction line shape synthesized from diffracted power increments $A, B, C, -$.

increments P , are adjusted by successive approximations such that the sum of the ordinates at any point along the abscissae gives the best fit to the experimental intensity curve. In Fig. 5 the sum of these increments taken at intervals of $2\frac{1}{2}$ percent in composition are seen to be in good agreement with the observed experimental curve. The diffracted power increments are simply the areas under the separate curves labeled $A, B, C, -$, and,

$$P_i = A_i (\pi/a)^{\frac{1}{2}}. \quad (2)$$

As we shall be interested only in the ratio of diffracted power increments from layers having similar a coefficients, the power increments are simply the amplitude factors A_i .

A considerably simpler method of synthesis is illustrated in Fig. 6. In this, a step function is fitted to the experimental curve and the areas of blocks $A, B, C, -$, represent increments of power scattered by the corresponding layers of Fig. 3. As the 5 percent breadth of these blocks is considerably less than the true line breadth, little significance can be attached to blocks at the extremities of the curve, e.g., H and I of Fig. 6. A first-order correction can be made however by fitting blocks to the zero-hour curve, Fig. 4, and using the results of a similar analysis on this as a subtractive correction to be applied to subsequent curves. These two methods of synthesis will be referred to as the error function method and the step function method of fitting.

The total diffracted power which is recorded on the photographic film is

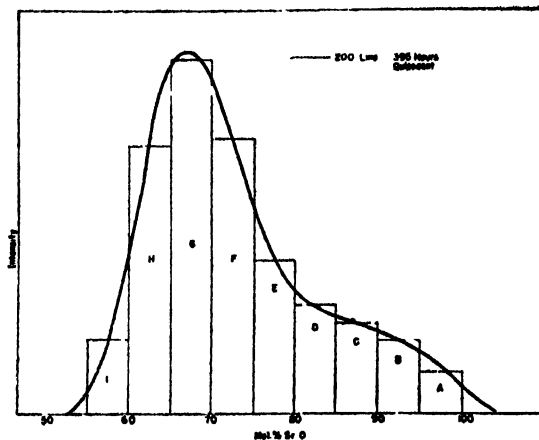


FIG. 6. 200 diffraction line shape synthesized from diffracted power increments $A, B, C, -$.

$$P_{\text{total}} = \sum_1^n P_i(\theta), \quad (3)$$

where

$$P_i(\theta) = A_i(\theta) P_i'(\theta), \quad (4)$$

and $P_i(\theta)$ is the total power arriving at the film, scattered by layer i into the diffraction cone of semi-apex angle 2θ ; $P_i'(\theta)$ represents a similar increment of scattered power neglecting absorption and $A_i(\theta)$ is an appropriate correction for absorption in both the incident and diffracted beams. The form of this correction has been discussed¹² for the cylindrical sample case of a thin layer of unknown thickness overlying a second layer having a considerably greater thickness. The thickness of this surface layer can be computed if the ratio of power diffracted from the two layers is known.

In the present treatment, increments of effective scattered power P_i , are treated in turn starting with P_A from the surface layer. Making use of the relationship¹² between the surface layer thickness and the scattered power ratio, there results,

$$\frac{P_A}{P_{\text{total}} - P_A} = \frac{N_A^2 (F_A^2 / \mu_A)}{N_K^2 (F_K^2 / \mu_K)} \times \frac{\int_0^\infty \frac{1}{y} (1 - \exp \{-\mu_A l_A y\}) dy}{\int_0^\infty \frac{1}{y} \exp \{-\mu_A l_A y\} dy}, \quad (5)$$

¹² A. Eisenstein, *Bull. Am. Phys. Soc.* 21, No. 1, A-12, (1946); *Phys. Rev.* 69, 252 (1946).

where

N = number of unit cells per unit volume,

F = molecular structure factor,

μ = linear absorption coefficient,

$\gamma = (1/\sin \alpha) + (1/\sin [2\theta - \alpha])$,

2θ = scattering angle,

α = angle of incidence between primary beam and the tangent to the cylindrical surface, and

l_A = thickness of layer A .

The subscript A denotes properties of the surface layer and K represents an average of the properties of all underlying layers. A numerical evaluation was made¹² of the integrals of (5), as a function of $\mu_A l$ for the particular values of 2θ which are of interest in this analysis. Designating the ratio of integrals as S_1/S_2 , then,

$$\frac{S_1}{S_2} = \frac{P_A}{P_{\text{total}} - P_A} \frac{N_K^2(F_K^2/\mu_K)}{N_A^2(F_A^2/\mu_A)} \quad (6)$$

A comparison of the power increment P_A from layer A with the power diffracted by all underlying layers, $P_{\text{total}} - P_A$, allows a determination of the layer thickness l_A . The diffracted powers $P_B, P_C, P_D, \dots$, can be lumped together as in (5) and (6) only if their $N^2(F^2/\mu)$ values are similar. As shown in Fig. 7b, this quantity varies by as much as 60 percent in the composition range 50 percent–100 percent SrO; however an error considerably less than this is introduced by selecting a weighted average value of $N^2(F^2/\mu)$.

Similarly P_B is compared with $P_{\text{total}} - P_A - P_B$ to obtain l_B or, as is usually done, P_A and P_B are lumped together ($P_A + P_B$) and with an average $N^2(F^2/\mu)$ are compared with $P_{\text{total}} - P_A - P_B$ to obtain $(l_A + l_B)$. This process is repeated until the thickness of each layer has been evaluated.

TABLE I. Monode cathodes coated with C51-2 and operated quiescent.

Base Metal	Coating temperature °C	Hours	Bulk % SrO	Surf.
Gr. A Ni	800	300	51 ± 2	none
Gr. A Ni	800	600	54 ± 2	thin
Gr. A Ni	875	50	49 ± 5	med. thin
Gr. A Ni	875	75	52 ± 5	med. thick
Gr. A Ni	875	75	50 ± 2	thick
Gr. A Ni	875	300	56 ± 2	med. thick
Gr. A Ni	875	600	62 ± 2	med. thick
Gr. A Ni	875	1200	62 ± 1	very thick
N-81 Ni	875	280	52 ± 1	med. thin
N-81 Ni	875	280	51 ± 2	med. thin
N-81 Ni	875	600	56 ± 3	med. thin
N-81 Ni	875	1200	58 ± 1	medium

Finally $l_A, l_A + l_B, l_A + l_B + l_C, \dots$ are plotted as a function of composition to obtain the composition distribution curves.

RESULTS

An analysis of the diffraction patterns of the 16 cathodes of Group 1 which had reached the "end of life" showed that with one exception all oxide coatings had lost nearly all of the initial BaO content. These 15 coatings whose life test periods varied from 500 to 1000 hours had an average final bulk composition of 95 ± 4 percent SrO. The one exception was a cathode prepared on a 5 percent Si-nickel base which had reached the "end of life" in 650 hours at 800°C under d.c. life test conditions. In this cathode a bulk composition of 52 percent SrO was found as well as a surface layer containing up to 95 percent SrO.

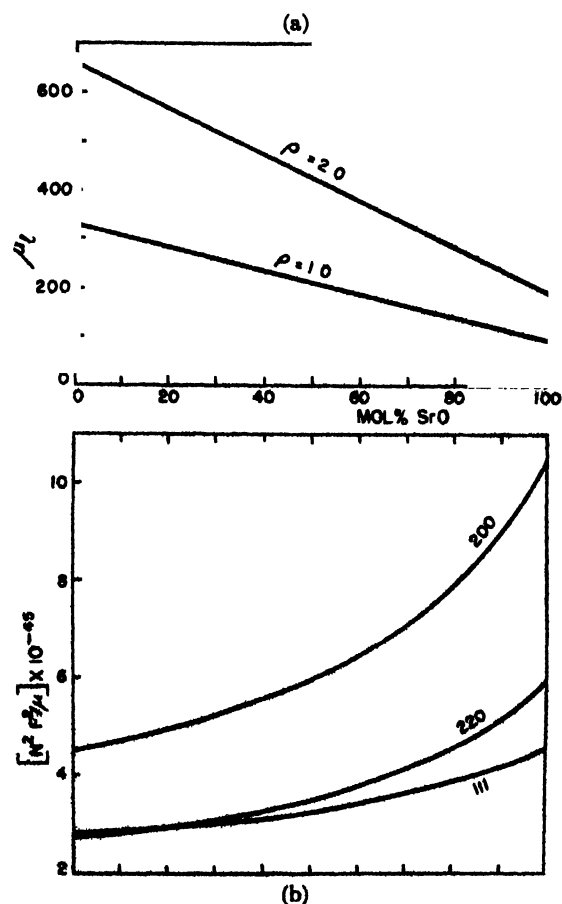


FIG. 7. (a) Computed linear absorption coefficient vs. composition for apparent densities ρ , of 1.0 and 2.0 g/cm³, and (b) computed N^2F^2/μ vs. composition for 111, 200, and 220 diffraction lines.

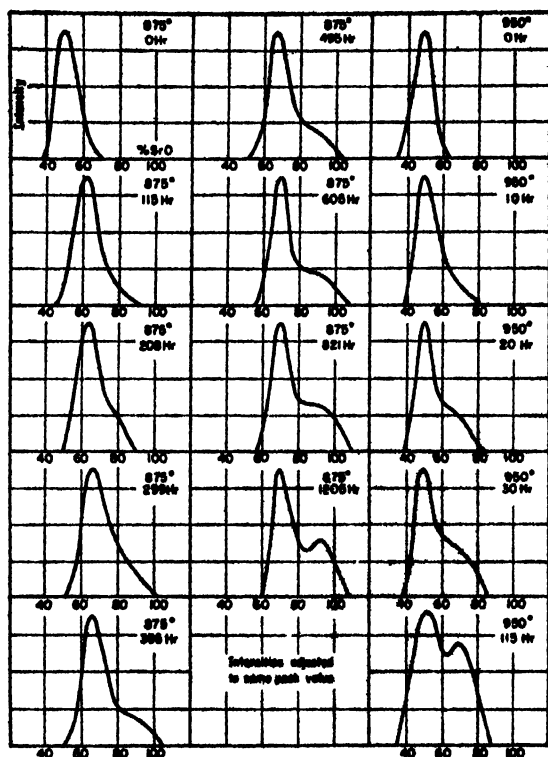


FIG. 8. 200 diffraction lines from cathodes 1 and 2 of Table II.

TABLE II. Diode cathode coated with C51-2 and operated quiescent.

Grade A nickel base (1) 875°C— pumped		(2) 870°C— pumped		Vacuum fired Grade A Ni base (3) 950°C— pumped		(4) 875°C— sealed off	
Hours	Bulk % SrO	Hours	Bulk % SrO	Hours	Bulk % SrO	Hours	Bulk % SrO
0	50	0	50	0	50	0	50
13	50±1	27	50	10	51±1	34	47±1
33	54±1	115	61±3	20	50±2	73	50±1
64	59±3	260	64±1	30	54±4	147	48±2
115	62±4			40	52±1		
208	65±3			50	52±1		
299	67±3			72	52±1		
395	67±3			115	49±2		
495	67±4						
605	69±4						
821	68±2						
1205	70±1						

Table I shows the results obtained from an analysis of the diffraction patterns taken of the cathodes from the twelve monodes of Group 2 which were life tested in a quiescent state. In those patterns which showed a marked asymmetry in the shape of the diffraction lines, measurements of the peak positions were used to determine the bulk composition. The indicated $\pm$ error represents the average deviation of the measured 111, 200, and 220 lines of the diffrac-

tion pattern and does not represent the absolute accuracy of the determination.

In general, it appears that the bulk loss of BaO as well as the surface loss is slightly less for the cathodes prepared on the purer nickel base. This is consistent with the presence of larger amounts of the reducing impurities Mn, Mg, and Si in the grade A nickel than in N-81 which were detected in a spectroscopic analysis. The effect of the operating temperature on the bulk loss of BaO is indicated by comparing the molar compositions of the grade A cathodes operated 600 hours at 800°C and at 875°C. A thicker BaO deficient surface layer is found on the cathodes operated at 875°C than on those at 800°C.

In the experiment designed to investigate the effect of a d.c. emission current density of 0.5 amp./cm² on the loss of BaO, Group 3, the following results were obtained. In 40 hours at a temperature in excess¹³ of 950°C, a thick surface layer was observed on all of the seven cathodes. The four which were operated quiescent had a final average bulk composition of 63±4 percent SrO and the three life tested at 0.5 amp./cm² showed a bulk composition of 61±3 percent SrO. It is evident that the effect of the d.c. emission current is less than the experimental error involved in the composition determination. The error introduced into both the bulk and surface layer determinations could be reduced by using a lower operation temperature thus avoiding excessive crystal growth in the oxide coating and the resulting "graininess" in the diffraction patterns.

Time changes which occur in the bulk composition of the oxide coating are shown in Table II and were obtained from diffraction patterns of cathodes operated in the pumped and sealed off beryllium windowed diodes, Group 4. The effect of the base metal on the bulk loss of BaO is seen by comparing the three diodes operated at about 875°C. The vacuum fired grade A base cathode shows no loss of BaO in 147 hours at which time the regular grade A cathodes have increased to over 60 percent SrO. The second vacuum fired grade A base cathode shows no

¹³ All cathodes were initially set at 950°C but line voltage fluctuations in an overnight life test caused all cathode temperatures to increase by an unknown amount.

detectable BaO bulk loss in 115 hours at 950°C. All four cathode sleeves were chemically cleaned and hydrogen fired prior to coating. Two of these were heated to 1050°C for 30 minutes in the additional vacuum firing treatment. The effect of this process is to evaporate possible reducing agents from the surface of the base metal. A spectroscopic analysis of the evaporated material shows it to contain major quantities of Mg, Mn, Cu, and Ni. All cathodes of this group showed asymmetrical diffraction lines indicating some surface loss of BaO as well. Cathodes 1 and 2 of Table II, Group 4, showed a more rapid loss of BaO than similar cathodes of Table I, Group 2. No completely satisfactory explanation was found for this difference.

The effect of time and temperature on the shape of the 200 diffraction line from initially equal molar BaSr(O) may be seen in Fig. 8. Microphotometer tracings converted to units of intensity and adjusted to the same maximum value are shown plotted as a function of the molar percent SrO composition. The grade A nickel base cathode operated at 875°C shows a progressive asymmetry change with time as well as a shift in the position of the peak (see Table II). The vacuum fired grade A nickel base cathode at 950°C shows only the progressive asymmetry which develops at a rapid rate.

By means of the analysis method previously described, the intensity *vs.* composition functions of Fig. 8 have been transposed to the composition *vs.* depth functions of Fig. 9. The error function method of synthesizing the 875°C curves results in the composition distributions of Fig. 9a. In 115 hours a layer containing less than 10 percent BaO is found to cover the surface to a depth of 10^{-5} cm while in 1205 hours this layer has reached a thickness of 3×10^{-4} cm. At this period a layer of pure SrO covers the surface to a depth of about 5×10^{-5} cm. The simpler step function method of analysis when applied to the same curves gives the distributions of Fig. 9b. Applying the zero hour, fortuitous curve as a subtractive correction to each of the other curves brings the results into fair agreement with the distributions obtained using the more laborious error function method. An analysis of the 950°C intensity curves using the step function method may be seen in Fig. 9c. The bulk composition

remains equal molar as the BaO deficient surface layers take form.

As a check on the validity of this method of composition determination, analyses were per-

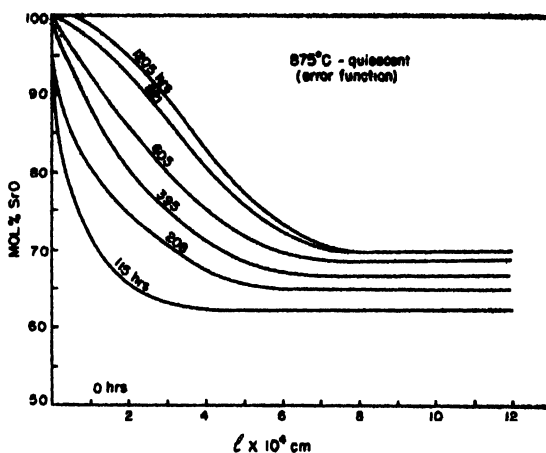


FIG. 9a. Composition *vs.* depth function, determined using error function method of fitting, 875°C

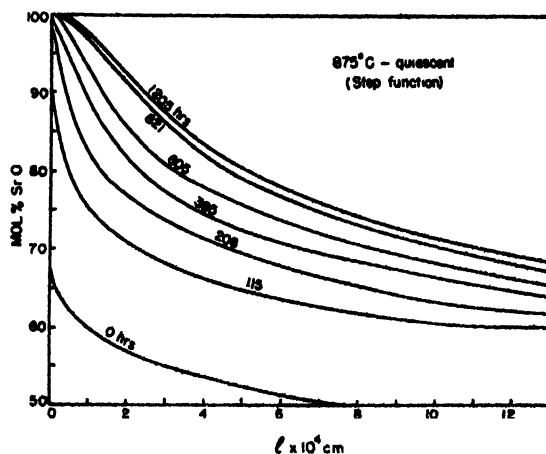


FIG. 9b. Composition *vs.* depth function, step function method of fitting, 875°C.

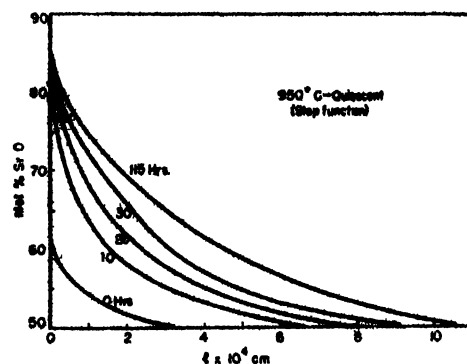


FIG. 9c. Composition *vs.* depth function, step function method of fitting, 950°C.

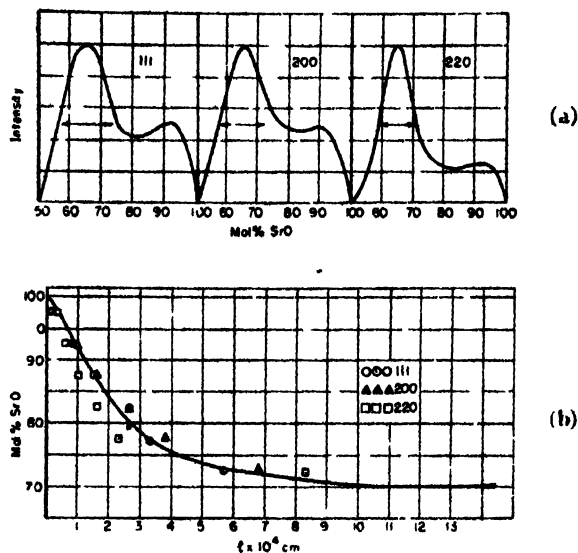


FIG. 10. (a) 111, 200, and 220 diffraction lines from cathode after 1225 hours quiescent at 875°C. (b) Composition vs. depth functions computed from these lines.

formed on three different diffraction lines of the same pattern. Figure 10a shows the intensity vs. composition traces of the 111, 200, and 220 diffraction lines from a cathode, initially equal molar BaSr(O), after 1225 hours at 875°C, quiescent. The marked effect of the scattering angle and the particular hkl reflection on the line shape is readily evident. Arrows at the half-maximum position indicate the true line breadth extrapolated from the measured breadth of nearby¹⁴ large crystal tungsten lines. The principal peak of these curves corresponds therefore to a single homogeneous phase in accord with the initial assumption regarding this structure. Analyses performed on these three curves using the step function method result in the distributions of Fig. 10b. The agreement between the computed distributions indicates the self-consistency to be expected from this method of analysis using different diffraction lines. This agreement serves also as a reasonable confirmation of the assumed monotonic distribution function. Actually, the BaO content of the coating may be very low¹⁴ in the oxide layers adjacent to the base metal because of the chemical affinity of Ba and certain base metal components.¹⁵ In general, these layers contribute very

little to the observed diffracted intensity as evidenced by the weak scattering from the adjacent base metal.

DISCUSSION

The loss of the BaO component of an oxide cathode coating initially equal molar (BaSr)O manifests itself as (1) a surface deficiency and (2) a bulk deficiency. It appears that, given sufficient time, all cathodes will develop a surface deficiency which is detectable using x-ray techniques. In this time interval the bulk loss may or may not be sufficient to permit its detection, depending upon the base metal used and the operating temperature (see Table I). The loss of BaO may occur as an evaporation of Ba or BaO. In either case the loss is presumably from the surface; hence, unless the temperature is sufficiently high to permit rapid diffusion of BaO or SrO, changes in the surface composition should be very sensitive to this evaporation loss. The results of Table II would seem to indicate that the removal of certain surface constituents from the base metal by the vacuum firing treatment eliminates the bulk loss without seriously impairing the loss from the surface. A comparison of the cathodes of Table I made on grade A nickel and run at 875°C with those made on the purer nickel, N-81, shows the effect of the base metal on the bulk loss and a lesser effect on the surface loss. This would seem to indicate that the bulk loss arises largely from a chemical reaction between the coating and reducing agents in the base metal, giving rise to free barium which diffuses to the surface and is lost through evaporation. Although a surface deficiency is found in cathodes made on the pure nickel, the observation that it is less than in cathodes made on grade A nickel indicates that at least part of the surface loss in the grade A cathode is caused by evaporation of Ba. Reducing impurities present initially in the coating may be responsible for this surface loss. The exact mechanism of the respective surface and bulk losses cannot be specified completely until more information is available regarding the diffusion rates of Ba, Sr, BaO, and SrO at these temperatures as well as the thermochemistry of the reactions involved. It is interesting to note that the surface deficiencies extend to depths of about 10^{-4} cm, or

¹⁴ L. A. Wooten, *Bull. Am. Phys. Soc.* 21, D-3 (1945).

¹⁵ A. Fineman and A. Eisenstein, "Studies of the interface of oxide coated cathodes," *J. App. Phys.* 17, 663 (1946).

roughly 8.1 of the total coating thickness. In the case of cathodes "flashed" at a high temperature to promote crystal growth,⁸ the observed composition variation probably occurs in single crystals on the cathode surface. It should be noted that an experimental upper limit exists on the composition vs. depth function that can be determined using this method. This is reached when scattering from the surface layers obscures the prominent peak from the bulk composition. This condition is nearly reached in the 115 hour, 950°C curve of Fig. 8. A lower limit of about 10^{-3} cm is indicated.

The effect of changes in the BaO-SrO composition during life on the thermionic emission properties of the cathode remains problematical. No direct evidence has yet been obtained to prove the existence of such an effect although several indirect experimental observations would seem to substantiate it. The experiments of Benjamin and Rooksby⁴ using d.c. thermionic emission cathodes, and Widell,¹⁶ using pulsed cathodes, indicate an optimum coating composition.

¹⁶ E. G. Widel, Bull. Am. Phys. Soc. 21, E2 (1946).

tion of near equal molar proportions. However, through the use of a high current density aging treatment, Fineman¹⁷ has been able to obtain a pulsed emission at 850°C to 900°C of 30 amp./cm² from a pure SrO coating and 60 amp./cm² from pure BaO, the latter value being in substantial agreement with that reported¹⁰ from (BaSr)O. Recent measurements of Nishibori and Kawamura¹⁸ give for the thermionic work functions of BaO and (BaSr)O about 1.0 ev and SrO 1.5 ev and for the conductivity activation energies (BaSr)O, 0.7 ev and SrO, 1.2 ev. The presence of a surface layer, deficient in BaO, may have an electronic conductivity somewhat lower than that of the bulk and, under certain conditions of operation, actually limit the thermionic emission.

The author is indebted to members of the Cathode Research Group for valuable assistance during the course of this investigation, particularly to Mr. A. Fineman who contributed most of the standard diode cathodes used in this study.

¹⁷ A. Fineman, private communication.

¹⁸ E. Nishibori and H. Kawamura, Proc. Phys. Math. Soc. Japan 22, 378 (1940).

Studies of the Interface of Oxide Coated Cathodes*

A. FINEMAN AND A. EISENSTEIN

Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

X-ray diffraction patterns are presented as evidence for the existence of a crystalline "interface" compound between the base metal and the coating of the oxide coated cathode. Measurements of the coating resistance to microsecond pulse currents, made with imbedded probes, indicate the presence of an anomalous "interface" resistance between the metal and coating. The relative magnitude of the coating and interface resistances are shown as a function of current for various operating temperatures.

I. INTRODUCTION

WHEN a system consisting of a metal coated with alkaline earth carbonates is heated in a vacuum, some combination of the metal and alkaline earth element takes place. The nature of the interface compound formed in the case of the oxide coated cathode has been postulated by various workers¹ to be of the Type ABO_n , where

* This paper is based on work done for the Office of Scientific Research and Development under contract OEMar-262 with the Radiation Laboratory, Massachusetts Institute of Technology.

¹ H. D. Arnold, Phys. Rev. 16, 70 (1920). A. Gehrts, Zeits. f. tech. Physik 11, 244 (1930). L. A. Wooten, Bell Telephone Laboratory, private communication.

A is the alkaline earth element and B, the base metal (e.g., $BaPtO_n$, $BaNiO_n$, $BaTiO_n$). Arnold states that $BaPtO_3$ was identified on a Pt wire cathode but gives no indication of the method by which it was identified. The other estimates were based upon chemical or spectrographic determinations of the presence of the base metal in the coating layers close to the metal. Direct evidence, in the form of x-ray diffraction patterns, is presented in this paper to show that specific crystalline compounds are formed at the metal-coating interface of oxide-coated cathodes. The coating resistance to pulsed microsecond

currents is shown to be anomalous in the region of the coating close to the metal. This anomalous resistance, which is termed interface resistance, greatly exceeds the coating resistance in most of the cases examined.

II. EXPERIMENTAL METHODS

A. X-Ray Diffraction Studies

The method used in obtaining x-ray diffraction patterns of oxide coated cylindrical cathodes has been described in detail.² The vacuum diode containing the cathode to be examined was operated in an atmosphere of dry nitrogen because of the chemical instability of certain of the

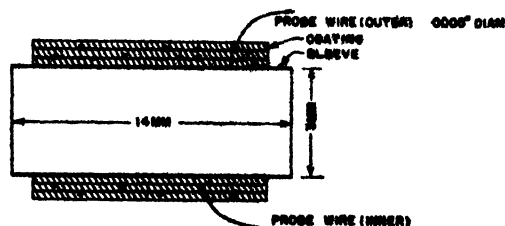


FIG. 1. Construction of double probe cathode.

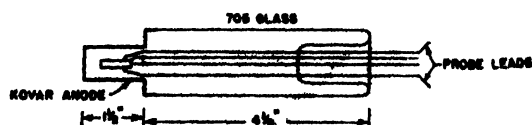


FIG. 2. Experimental diode

interface compounds. The cathode coating was scraped off until the interface became exposed, and the cathode was coated with a protective wax. The specimen was mounted in a 4.70-cm radius x-ray camera and the exposure made using filtered $\text{CuK}\alpha$ -radiation. In some cases only half of the cathode was dipped into the wax, and patterns were then taken of the waxed and non-waxed portions in order to check the stability of the interface compound to oxidation by air or water vapor.

B. Resistance Measurements

Coating and interface resistance measurements were made by the use of embedded probes.³ Figure 1 illustrates the construction of the double probe cathode. The cylindrical sleeve was weighed, mounted in a rotatable chuck, sprayed with carbonate coating, and re-weighed. Several turns of probe wire were tightly wound around the coating and a second layer of carbonate coating was sprayed on. The cathode was again weighed and the process repeated with a second probe wire.

Figure 2 illustrates the test diode in which measurements were made, and Fig. 3 shows the electrical circuit. Determinations of pulsed anode voltage V_A , microsecond pulse current I , and probe pulse potentials (V_{p1} , V_{p2}) were made as indicated. The total resistance from the sleeve

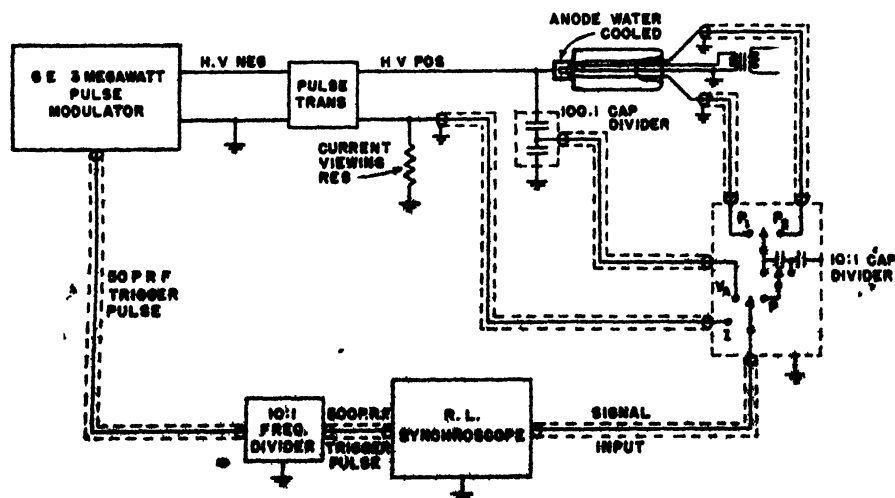


FIG. 3. Measuring circuit.

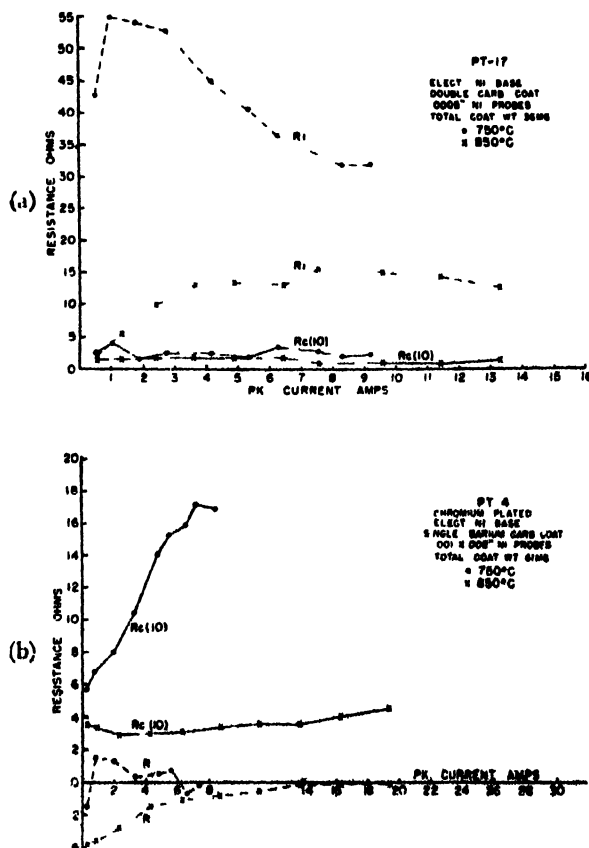
² A. S. Eisenstein, "A study of oxide cathodes by x-ray diffraction methods" (preceding article, this issue).

³ E. Nishihara and H. Kawamura, Proc. Phys. Math. Soc. Japan 22, 378 (1940). W. E. Danforth, Bartol Research Foundation, private communication.

to each probe was obtained by dividing the corresponding probe potential by the current. The difference between these two resistances was taken as the resistance of the amount of coating lying between the probes. The resistance of the coating between the sleeve and first probe was then computed on the basis of simple weight proportion, and the subtraction of this amount of coating resistance from the total resistance between sleeve and first probe gave the quantity which was termed interface resistance, R_i . The quantity $R_{c(10)}$ was defined as the resistance of 10 mg of coating. All cathodes were prepared with an area of 1 sq. cm.

III. RESULTS

Interface patterns were found on all coating, base-metal systems examined. These included base metals of copper, pure and Grade "A" nickels, chromium plated nickel, and alloys of Si-nickel and Cr-nickel.



A. Grade "A" and Pure Nickel Base Metals

The interfaces formed on Grade "A" nickel were light grey in color and usually too thin to give analyzable patterns with the x-ray technique used. In the case of the pure nickels a consistent pattern was obtained. However, because of the difficulty of obtaining good diffraction patterns of the interfaces on the pure and Grade "A" nickels, and since the interfaces observed on

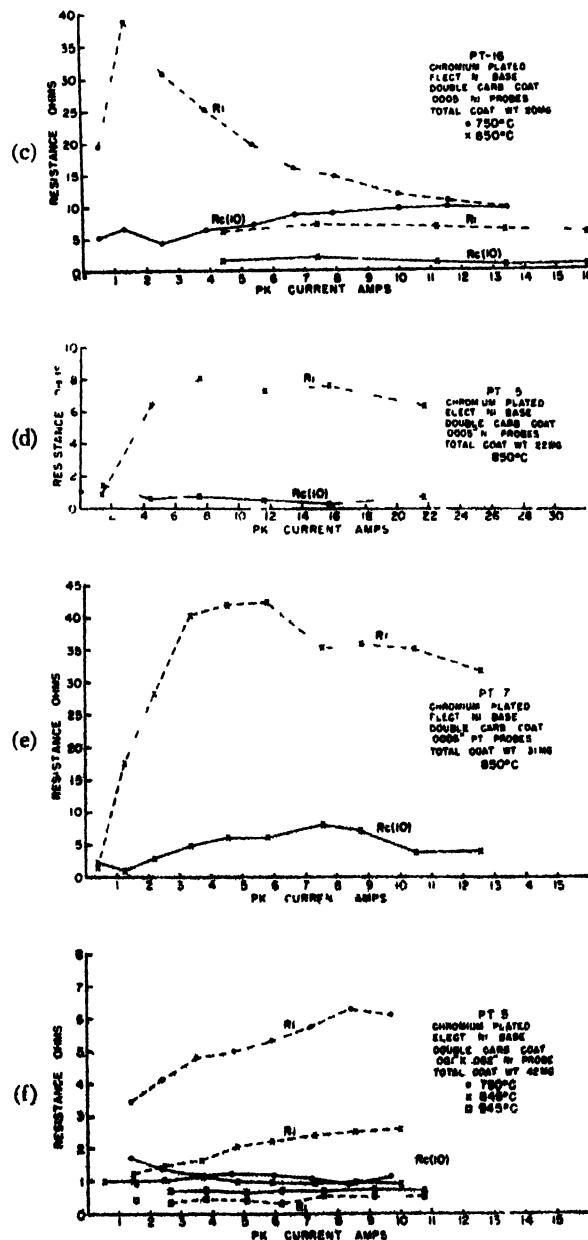


FIG. 4. Coating resistance $R_{c(10)}$ and interface resistance R_i versus peak current I .

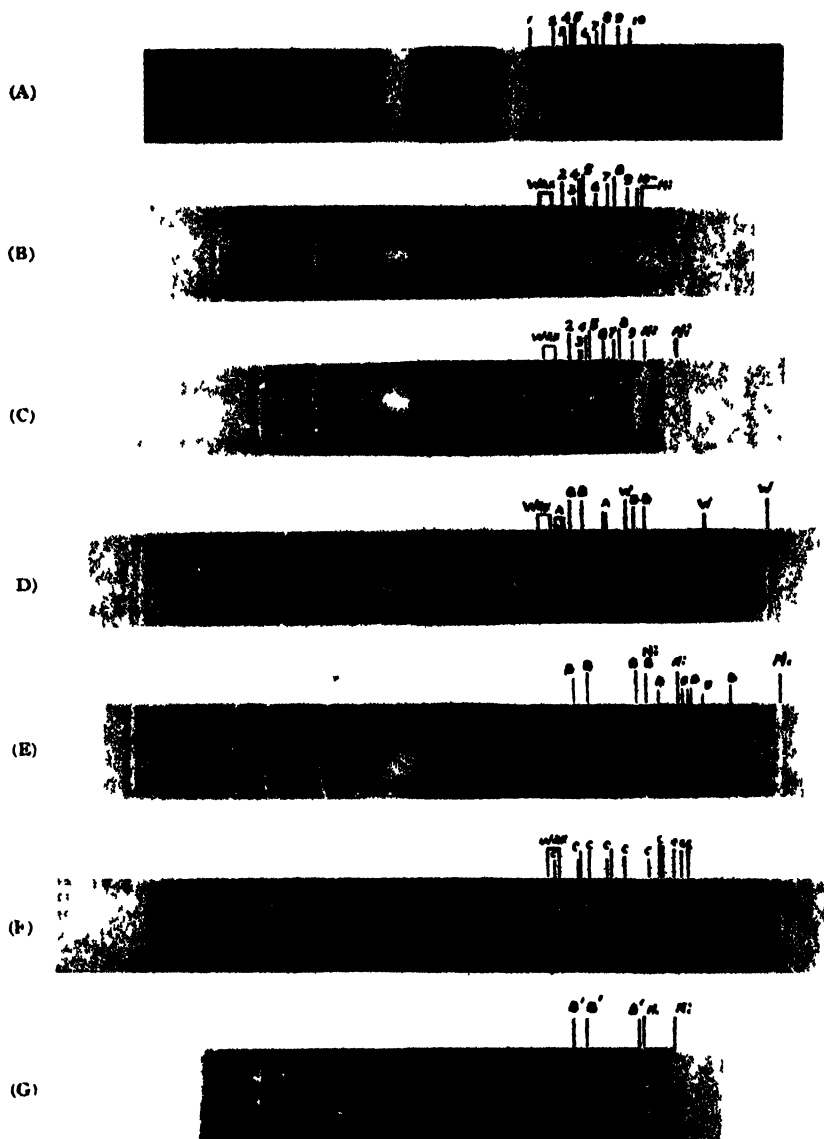


FIG. 5. X-ray diffraction patterns of interfaces:

- (A) Synthetic "BaSiO₃" prepared by heating BaCO₃ and SiO₂ in air, 1 hour at 1000°C (0.5 mm sample diameter);
- (B) "BaSiO₃" found at interface of cathode made from (BaSr)CO₃ on 5 percent Si-Ni base;
- (C) Compound found at interface of cathode made from SrCO₃ on 5 percent Si-Ni base;
- (D) Interface compounds "A" and "B" from cathode of BaCO₃ on Cr plated Ni;
- (E) Interface compound "B" from cathode of either BaCO₃ or (BaSr)CO₃ on Cr plated Ni;
- (F) Interface compound "C" from cathode of BaCO₃ on Cr plated Ni;
- (G) Interface compound from cathode of SrCO₃ on Cr plated Ni.

cathodes made of Cr and Si nickel alloys⁴ were consistent and more easily identified, these latter systems were chosen in preference to Grade "A" and pure nickels for the more extensive studies of the interface. Figure 4a shows a plot of the interface and coating resistances as a function of microsecond pulse current for a pure nickel cathode coated with (BaSr)CO₃.⁵ The data given are for pyrometer temperatures of 750°C and 850°C. It is noted that the coating resistance is

essentially ohmic, and that the interface resistance which greatly exceeds the coating resistance has a decided maximum which shifts to lower currents as the temperature is decreased. The dependence of resistance upon temperature is seen to be more pronounced in the case of the interface than in that of the coating.

B. Interface Compounds Formed on Si-Ni Alloy Base Metal

Cathodes prepared from double carbonate (BaSr)CO₃ on 2 percent or 5 percent SiNi alloy showed an interface pattern identifiable with

⁴ All pure nickels and alloys used in this work were obtained through the kind cooperation of Mr. E. M. Wise of the International Nickel Company.

⁵ Coating C51-2 obtained from the Raytheon Manufacturing Company, Newton, Massachusetts.

that of synthetically prepared BaSiO_3 ⁶ (see Figs. 5A and 5B). The d values of the interface and synthetic compounds are compared in Table I. The synthetic BaSiO_3 was white while the interface " BaSiO_3 " showed varying shades of grey, probably because of the interspersions of finely divided nickel. Although $(\text{BaSr})\text{CO}_3$ coated cathodes at different states of activity and at d.c. and microsecond pulsed life from 0 to 500 hours were examined for interface patterns, the " BaSiO_3 " pattern alone was observed in every case. This interface compound was stable to both air or water exposure.

One cathode was prepared with pure SrCO_3 ⁷ substituted for $(\text{BaSr})\text{CO}_3$ on the 5 percent Si-Ni alloy. The interface compound here was similar in crystal structure to " BaSiO_3 " but had consistently smaller d values, caused probably by the direct replacement of Ba with Sr in the compound. See Fig. 5C. No resistance measurements were taken on these systems.

C. Interface Compounds Formed on Cr Plated Nickel

In the case of cathodes prepared of BaCO_3 on Cr plated pure nickel three separate interface compounds were identified. Their diffraction patterns did not correspond to any compounds of Ba and Cr listed in the ASTM x-ray diffraction card file. These compounds which are here arbitrarily designated as " A ," " B ," and " C " made their appearance at various levels of cathode activity and life (see Figs. 5D, E, and F).⁸ The interface compounds were all clearly visible and colored. Those interfaces in which only " B " was detected showed homogeneous light or dark browns and in some cases brown mottled with green. " A " was never found alone, and " C " in

the one case in which it was found alone showed dark brown. Both " A " and " C " were converted to " B " upon exposure to air. " B " was stable to both air and water at room temperature and was not affected by heating in air at 850°C for 15 minutes. Analyses of the " B " compound with the Davey-Hull charts indicated a possible close packed hexagonal structure with the d values given in Table II.

Cathodes prepared with SrCO_3 on Cr plated nickel showed an interface compound having a crystal structure similar to that of " B " but with somewhat smaller lattice constants. (See Fig. 5G.)

In the case of cathodes prepared of $(\text{BaSr})\text{CO}_3$ on Cr plated nickel, only the " B " compound was identified. It is to be noted that as in the Si series, where both Ba and Sr were present in the coating, the Ba, alone, took part in the formation

TABLE I. Observed d values and line intensities of the interface and synthetic BaSiO_3 compounds.

Interface		Synthetic BaSiO_3	
d	I	d	I
4.14	M	4.11	M
3.38	S	3.36	S
3.11	M	4.09	MW
2.97	VS	3.97	VS
2.89	VS	2.87	VS
2.67	VW	2.66	W
2.47	W	2.48	VW
2.38	MW	2.38	M
2.21	MW	2.20	M
2.06	M	2.07	M
1.865	W	1.875	MW
1.82	W		
1.73	W	1.725	M
1.67	M	1.66	W
1.43	W		
1.41	W		

TABLE II. Compound " B " indices from Davy-Hull chart for simple hexagonal system.

Computed values		Compound " B "	
Indices	d	d	I
101	3.19	3.21	VS
110	2.89	2.91	S
200	2.50	2.50	VW
201	2.14	2.16	S
002	2.08	2.08	M
102	1.92	1.92	M
121	1.72	1.725	M
112	1.69	1.690	M
202	1.60	1.615	MW
122	1.46	1.415	W
003	1.385	1.340	M

Computed d values for $a_0 = 5.77$, $c/a = 0.72$.

⁶ Finely powdered SiO_2 and BaCO_3 mixed in bulk and heated in air or vacuum to 850° to 1000°C . See W. Jander, *Zeits. anorg. allgen. Chemie* 163, 8 (1927).

⁷ Contained less than 0.01 percent Ba. Prepared by C. D. Prater, Bartol Research Foundation, Swarthmore, Pennsylvania.

⁸ Bulk synthesis of the barium-chromium interface compounds, A , B , and C , was attempted by heating the following combinations: (a) 150-mesh chromium powder and barium carbonate, (b) chromium powder and barium chromate, and (c) chromium oxide (Cr_2O_3) and barium carbonate. Samples were heated in vacuum and in the air, at temperatures of 850°C and 1050°C , for times varying between 30 minutes to several hours. None of the interface compounds, A , B , or C , were observed in any case.

of the interface compound. The "B" compound was also identified in cathodes prepared with $(\text{BaSr})\text{CO}_3$ on 5 percent CrNi alloy.

Plots of $R_{c(10)}$ and R_i vs. I for a cathode composed of BaCO_3 on Cr plated nickel are shown in Fig. 4b. Here, the coating resistance exceeded the interface resistance, and showed more dependence on temperature. Figures 4c, d, e, and f show resistance vs. current curves for four cathodes made of $(\text{BaSr})\text{CO}_3$ on Cr plated nickel. Each of these tubes showed the Type "B" interface. It may be noted that usually the interface exceeds the coating resistance and is more temperature dependent than is the coating resistance in the range studied. Figure 4f shows, however, that the interface resistance may drop below the coating resistance at elevated temperatures. Variations in coating resistance from tube to tube are seen to be much smaller than corresponding variations in interface resistance.⁹

IV. DISCUSSION

The plots of coating and interface resistance shown above suggest that the interface may be a

⁹ The R vs. I curves shown were all obtained with either nickel or platinum probe wires. Tubes were also made with probes of Rh-Pt alloy, pure Fe, pure W, and pure Cu. In these cases and also in the case of some of the nickel and platinum probe tubes the R vs. I curves were of the type shown in Fig. 6.

All tubes exhibiting this type of characteristic on either one or both probes were discarded for the following reason: impedance measurements at 1 megacycle, taken between sleeve and probe, showed that the tubes having a peaking R vs. I (Fig. 4) characteristic, demonstrated a shunt resistance which fell from megohms at room temperature to hundreds of ohms at 850°C , while the tubes showing a dropping R vs. I characteristic (Fig. 6) maintained a shunt resistance of megohms throughout this temperature range. The latter phenomenon was interpreted as meaning that in these cases the probe was electrically insulated from the coating.

factor which enters into the limitation of pulsed emission in oxide coated cathodes.

The interface resistance curves of Fig. 4 when plotted against the voltage drop across the interface (IR_i), are similar in shape and in variation with temperature to those exhibited by multiple

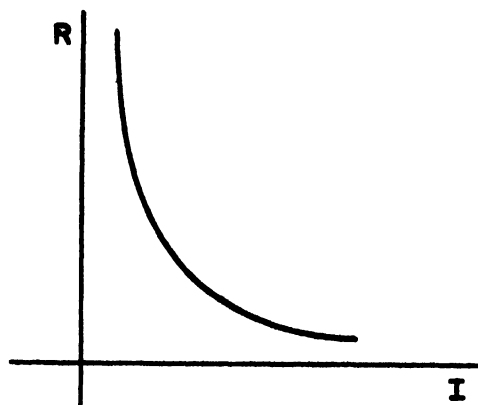


FIG. 6.

layer oxide rectifiers.¹⁰ Since, in addition to this fact the x-ray data show the presence of a definite crystalline compound between the coating and the base metal, it is suggested that the oxide coated cathode is dependent for its operation upon the same physical structure (metal + blocking layer + semiconductor) as is present in the oxide rectifier.

The authors wish to express their gratitude to Dr. E. A. Coomes and members of the Cathode Research Group for their helpful suggestions and invaluable aid throughout the course of this investigation.

¹⁰ N. F. Mott, Proc. Roy. Soc. A171, 127 (1939).

Midwest Research Institute—A New Scientific Research Institution

BY GEORGE E. ZIEGLER

Executive Scientist, Midwest Research Institute, Kansas City, Missouri

DURING the past year a new scientific research institution placing emphasis on research in physics and its closely related sciences has come into being in Kansas City, Missouri. The Midwest Research Institute a year ago was just the dream of Middle America's business men; now it is a going scientific institution with a staff of over ninety working on both research for industry and on regional research to benefit the people of Middle America as a whole. The Institute's research program is balanced between physics, with its closely associated engineering sciences, and chemistry, with its many branches. It is the purpose of this article to describe briefly the Midwest Research Institute and to illustrate how its well-coordinated physical research and measurements groups serve as the mainstay for many types of engineering and chemical research, as well as for industrial physics research.

In 1943 a group of public-minded business men saw the need for an outstanding scientific institution of broad base to aid in the region's forthcoming industrial development. They crystallized their vision by calling to a Midwest Research Council meeting leading research men from the Middle Western universities and colleges, as well as state geologists and business people familiar with the research needs and potentialities of the region. Following through on their first step this group of men, under the far-seeing guidance of Mr. Robert L. Mehornay and Mr. J. C. Nichols—business men and civic leaders of the Kansas City Area—contributed over \$500,000.00 in 1944 to serve as an initial operating fund for the new institution.

The soundness of the vision of this group of business men made it possible to attract experienced personnel to the Institute for both administrative and laboratory research work. At the present time this same group of business men, now organized as a group of trustees with a board of governors from their own number, has

embarked on a much larger fund raising campaign to enable the Institute to increase still further its facilities and to increase significantly its program of regional research so that Middle America will be assured of benefits that can result from research.

PLAN OF ORGANIZATION AND STAFF

The establishment of Midwest Research Institute by its Board of Governors and Trustees as a not-for-profit, scientific institute standing on its own feet gives it a unique opportunity for service. It has the advantage of being able to cooperate with all the universities and colleges throughout the area at the same time it renders research service to industry and government.



FIG. 1. High speed photographic study of chip formation with different cutting tools.



FIG. 2. 120,000-pound tensile testing machine in operation in the engineering mechanics laboratory.

This was demonstrated recently when it sponsored a seminar on high polymers. This two-day seminar was attended by the senior staff members of fourteen universities and twenty-five industrial concerns.

The Midwest Research Institute secured as its president Harold Vagtborg, whose pioneering spirit and vigorous organizational ability brought Armour Research Foundation, Chicago, to its outstanding position among research organizations. The basic plan of operation of the Midwest Research Institute organization provides for a full utilization of his stimulating leadership.

Frank H. Trimble, chairman of the Physics Section, received the Ph.D. degree from the University of Missouri for his work in x-ray diffraction before going to the Armour Research Foundation in Chicago. Here he first supervised extensive research in physical metallurgy and then served as co-chairman of physics research.

The chairmen of Agricultural Chemistry, Inorganic Chemistry, and Organic Chemistry Research are Charles L. Shrewsbury, formerly of Purdue University; George W. Ward, of the

National Bureau of Standards; and Carl M. Marberg, formerly of Interchemical Corporation. The Engineering Mechanics staff is under the chairmanship of Martin Goland, formerly of Curtiss-Wright Corporation Research Laboratories, with C. O. Dohrenwend, formerly Chairman of Engineering Mechanics Research at the Armour Research Foundation, as Research Consultant.

The present staff of ninety-two consists of fifty-two research workers, twenty of whom have the Ph.D. degree, seven the master's degree, and twenty-five the bachelor's degree. There are also sixteen laboratory and technical assistants, and twenty-four non-technical employees.

In its general plan of operation the Institute consists of a modern streamlined industrial research organization in which decisions concerning research methods and research expenditures are made by men as close as possible to the actual work. The research group leader on a particular project makes day by day decisions as to what research procedures shall be followed and what equipment procured. The chairman of a section thus can devote his major attention to "quality control" of the research and to looking after the sponsor's interests.

Since Midwest Research Institute is a public service, not-for-profit scientific institution it has heavy responsibilities to the community as a whole. The broad direction of the Institute's program to benefit the people of Middle America, as well as helping American industry solve its specific problems, is a major responsibility of the Institute's president, in addition to those normally incumbent on a chief executive officer.

The author, as Executive Scientist, has a unique set of responsibilities consisting of: (a) The coordination of the research of all sections so that each research problem will be pursued along the lines that currently appear to be most promising, (b) the development of cooperative working relations with the universities and colleges, and (c) the stimulation of professional advancement of the Institute's research staff.

INTERRELATED PHYSICAL TOOLS

Despite the difficulties of war procurement and reconversion delays, the Institute has been able to assemble during its first year a well-rounded

complement of physical tools which permit the Physics Section to carry out a broad based but integrated program of physics research for industry, and to render invaluable aid to metallurgy, engineering, and the various chemical researches.

In developing the Institute's physical facilities and staff, special care has been exercised so that each tool or technique may be used in its range of greatest usefulness. Value and usefulness of individual scientific tools is increased by their integration into planned group application. The plan at Midwest Research Institute is so complete that its description is worth consideration as an illustration of the variety of tools that can converge on a single problem.

For example, equipment to study the appearance of materials includes a latest model Bausch and Lomb research metallograph for opaque objects, or the surface condition of transparent objects; a Leitz polarizing petrographic microscope for studying transparent materials and for carrying out microchemical analyses; as well as a stereoscopic microscope for general studies of form. For those cases where the particle size is too small for ordinary optical equipment the new universal model RCA electron microscope has been installed and is kept in condition for immediate use.

If information is needed concerning the crystal structure of materials, it can be obtained through the use of the General Electric x-ray diffraction unit with its assortment of tubes and newest type cameras. The nature of surfaces can be analyzed by use of the electron microscope operating as an electron diffraction unit. This assemblage of optical, x-ray, and electron optics equipment makes it possible to solve many industrial problems directly, and these tools are indispensable aids to metallurgy and chemistry.

Another important aid to metallurgy and chemistry is a complete Dietert spectrographic analysis unit. Either qualitative or quantitative analyses can be made by the arc or spark method, and the spectrograph has been set up with its camera extending into its own dark room, so that the photographic work can be carried out without delay or inconvenience. The Dietert spectrograph has been supplemented by an ultraviolet Beckman spectrophotometer and a

Coleman spectrophotometer. These instruments have proved very useful in glass technology research, as well as for the usual chemical analyses which are so conveniently made by absorption measurements.

The Institute's equipment for physical metallurgy consists of a 30-kilowatt Tocco high frequency generator, a 2-kilowatt Westinghouse 15-megacycle generator, Rockwell and Brinell hardness testers, and a vapocarb heat treating unit, together with quenching equipment. Complete metallurgical specimen preparation, as well as the physical testing equipment which has been primarily secured for physics research in Engineering Mechanics, is also available.

Engineering mechanics laboratory measurements can be carried out for both static and dynamic states. For statics there are Baldwin-Southwark tensile machines of 120,000-pound and of 60,000-pound capacity. There is also a 10,000-pound Dillon unit which is useful for plastics and light alloys. For work in dynamics there are available special high frequency response bridge circuits and impact equipment.

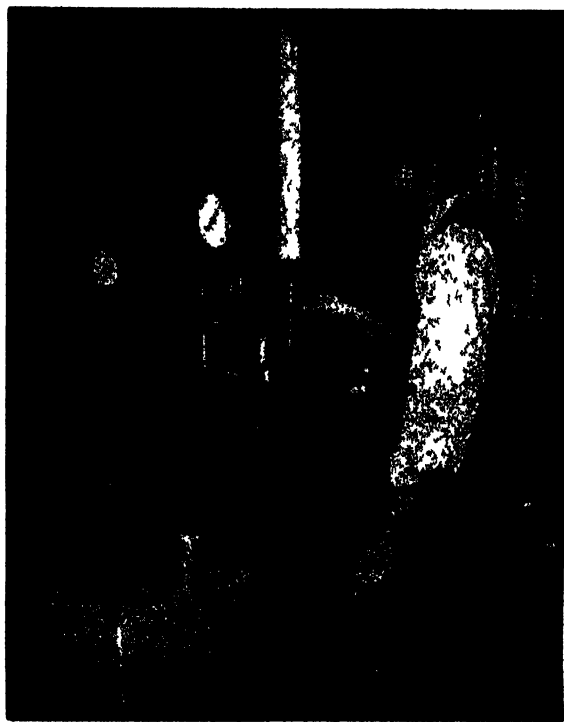


FIG. 3. Hardness testing equipment in metallurgical laboratory.

For certain special impact problems modifications of the Tinius Olsen metallurgical impact tester can be used.

One part of the Physics Section which is of special value to the entire Institute is a complete photographic laboratory and studio. Here photographs of any type can be made, either as part of research data or for reports. Regular studio production type printer, washer, and drier are installed. A liberal supply of all kinds of special photographic film and paper is kept on hand to be available for immediate use. The motion picture equipment includes an Eastman high speed 16-mm camera capable of taking up to 3000 frames per second. The Photographic Laboratory is equipped with an Ozalid machine for quickly making copies of drawings and reports.

The staff and equipment of the Electronics Laboratory have also proved of great value to all other sections of the Institute. The equipment now available includes an assortment of General Radio bridges and meters, as well as various types of oscillators, oscilloscopes, and a recording high frequency response galvanometer and the usual assortment of component parts for building up special electronic circuits.

A physics section in a broad based research organization is in a position to render more assistance to all of the other branches of research than is true of any other section, because of the fundamental nature of physics equipment and techniques. Nevertheless, physics research as such does receive invaluable aid from the other sections. At Midwest Research Institute there is available a complete, fully staffed analytical laboratory so that accurate information can be obtained on the chemical nature of material under investigation. Many times the solution of problems in physics has been made easier through consultation and advice from the chemistry staff concerning the availability of materials or concerning means of making the needed materials out of those available.

Of inestimable service to the Physics Department is the Institute's well-equipped Instrument and Machine Shop. In fact, the Physics Department, including Engineering Mechanics and Metallurgy, is one of the major customers of the shop.



FIG. 4. Projection densitometer being used for spectrographic analysis.

RESEARCH PROJECTS—REGIONAL AND INDUSTRIAL

One regional research project in physics has as its purpose the making available to the industry of Middle America as many physical techniques as possible. For example, one group is becoming proficient in the operation of the electron microscope, so that when an industry calls for some electron microscope photographs it will be possible to render not only a quick service, but an economical one. Similar procedures are being followed for all of the equipment mentioned in the description of facilities. This project is a particularly good one from the viewpoint of the professional standing of the men in the Physics Section because it provides an opportunity for publishable research. The physics staff also actively participates on other regional projects which consist of:

1. Relation of properties of wheat to the baking properties of flour.
2. Improvement of wood properties through chemical treatment.
3. Inexpensive surface hardening treatment for barn and feed lots.
4. Non-fuel uses of petroleum and natural gas.
5. Beneficiation of the marginal coals of the region.

The industrial research of the Institute is carried on in confidence with all patents being fully assigned to the sponsor. Nine industrial projects were completed during the first year and twenty-six are now in progress with a budget of approximately \$400,000 for the forthcoming year. Space in this article prohibits a detailed description of the projects now under way. This

has been included in the Institute's published annual report.

The whole program of the Midwest Research Institute is proving to be an excellent opportunity for physicists to demonstrate to the satisfaction of all that the arts, skills, and science of an experienced research physicist can properly enter the solution of every industrial research problem.

Here and There

New Appointments

Richard C. Darnell, consulting engineer on instrument design and application, has been appointed by the Department of State, under its cultural cooperation program, to serve in China as a specialist in scientific instruments and laboratory equipment. He will assist in the selection of modern instruments needed to replace those worn out or looted from universities and research organizations. He will remain in China for about six months.

Raymond T. Ellickson, assistant professor of physics at the Polytechnic Institute of Brooklyn, has accepted an appointment as associate professor of physics at Reed College, Portland, Oregon.

James W. McRae, electro-visual engineer for Bell Telephone Laboratories, has been appointed director of radio projects and television research for that organization. Ralph Bown, assistant director of research at Bell Telephone Laboratories since 1944, has been named director of research.

David S. Saxon joined the research staff of Philips Laboratories, Inc., on April 1, 1946, as an associate physicist and is in charge of the section on theoretical physics. Edmund S. Rittner joined Philips' research staff on May 15, 1946, as associate chemist. His work will be carried on in the photo-cell laboratory.

Brooklyn Polytechnic Symposia

A series of eight symposia on "Recent Progress in the Field of High Polymers" will be conducted at the Polytechnic Institute of Brooklyn, October, 1946 through May, 1947.

Awards

Raymond D. Mindlin, associate professor of civil engineering at Columbia University, has recently been awarded the Medal for Merit by order of President Truman for his wartime work in the development of the radio proximity fuse. Professor Mindlin was largely responsible for the mechanical design of the radio tubes in the proximity fuse.

For distinguished work in glass technology, the honorary degree of Doctor of Science was awarded to William C. Taylor, vice president and director of glass technology of Corning Glass Works, at the 110th annual commencement exercises held at Alfred University June 10.

Stevens Institute Fellowships

The Experimental Towing Tank, Stevens Institute of Technology, has announced the establishment of several fellowships for graduates of engineering colleges who are interested in learning the techniques of model tests of ship and flying-boat hulls as well as in acquiring experience in the broader principles of hydrodynamic research in which the Tank is engaged. For information and application forms address Professor John P. Fife, Director of Personnel, Experimental Towing Tank, Stevens Institute of Technology, Hoboken, New Jersey.

Liquids Research Begun at Stevens

Research on the properties of liquids has begun at Stevens Institute of Technology as the result of a grant-in-aid received recently by the college from The Research Corporation. The research concerns the study of liquids in the far ultra-violet section of the spectrum where the air is completely opaque, making it necessary to do the work in a very high vacuum. The liquids being used are mainly melted salts, and the research project will examine what happens internally when solids melt. Results of the experiments made possible by the \$4000 grant-in-aid will be published on completion of the research.

Training Program of the Clinton Laboratories

Object

The Clinton Laboratories, which are operated by the Monsanto Chemical Company under contract with the Manhattan District, are planning to conduct a combined training and research program during the fiscal year 1946-47. The purpose of this program is to aid in the dissemination of information, both fundamental and applied, concerning chain-reacting piles and their uses for both science and engineering. The facilities of the Laboratories will be made available to about 35 highly trained scientists and engineers from academic and industrial laboratories who have not had previous experience with piles. Those selected will have an opportunity to attend a

coordinated lecture program on the theory and operation of piles and to participate in the research and development program of the Laboratories. This program includes the following two broad fields:

- A. *Fundamental Research in the Fields of Nuclear Physics and Chemistry.* This research will involve both theoretical and experimental work.
- B. *Design of Piles.* This program has as its principal objective the extension of the technology of pile development into the important fields of peace-time application.

Organization of Laboratories

The key administrative staff of the Clinton Laboratories is as follows:

Research Director: Professor E. P. Wigner.
(On leave from Princeton University.)
Director of Physics Division: Dr. L. W. Nordheim.
Director of Chemistry Division: Dr. J. R. Coe.
Director of Technical Division: Dr. M. C. Leverett.
Director of Power Pile Division: Dr. C. R. McCullough.
Director of Training Program: Dr. Frederick Seitz.
(On part-time leave from the Carnegie Institute of Technology.)

The Research Divisions of the Laboratories have full-time staffs engaged in the activities of the Laboratories. These staffs include scientists and engineers who have been involved in nucleonics for a number of years. The participants of the training program are expected to work in close cooperation with the full-time staff in order that the Laboratories may function as a well-coordinated unit. Assignments will be made *primarily* on the basis of the interests of the participants; however, the interests of the Laboratories must also be kept in view.

Time

The training program began July 1, 1946, and will extend for the entire fiscal year. The formal lectures will not begin until about September 1, 1946. However, it is highly desirable that the participants arrive as soon after July 1 as possible in order to become familiar with the program of the Laboratories and to become adjusted to participation in it.

Nature of Lecture Program

The lectures will be given by a select group of physicists, chemists, and engineers. The major portion of these will be chosen from the Laboratories' staff, but a substantial fraction will be visiting scientists and engineers who have had extensive experience in the field of nuclear science and engineering. The topics of the lectures will be chosen from the following fields:

- A. Fundamental Nuclear Physics.
- B. Pile Science (Theory and Operation)
- C. Radiochemistry.
- D. Engineering Design of Piles.

Housing

The Clinton Laboratories will request living accommodations in Oak Ridge for the participants. These accommodations consist of individual houses, apartments, or dormitory rooms, depending upon individual needs, in the town of Oak Ridge, which is a fully equipped modern community. Transportation is available from Oak Ridge to the Clinton Laboratories. Requests for housing will be handled by Dr. Prescott Sandidge (Address: Box 1991, Knoxville, Tennessee). Moving expenses can be arranged.

Salaries

All participants will be transferred to the payroll of the Clinton Laboratories. Questions concerning salaries will be referred to Dr. Prescott Sandidge. In general a no-gain, no-loss policy will be adopted.

Application for Participation

Such applications should be made to Dr. Frederick Seitz, Department of Physics, Carnegie Institute of Technology, Pittsburgh, 13, Pennsylvania.

As stated above, questions relating to housing and salary will be referred to Dr. Sandidge.

Calendar of Meetings

August

- 19-20 Mathematical Association of America, Ithaca, New York
- 20-23 American Mathematical Society, Ithaca, New York

September

- 9-13 American Chemical Society, Chicago, Illinois
- 16-20 The Instrument Society of America, Pittsburgh, Pennsylvania
- 18-21 Illuminating Engineering Society, Quebec, Canada
- 19-21 American Physical Society, New York, New York
- 26-28 American Physical Society (Division of High Polymer Physics) and The Fiber Society, Charlottesville, Virginia (Joint Meeting)
- 30-October 3 American Society of Mechanical Engineers, Boston, Massachusetts

October

- 3-5 National Electronics Conference, Inc., Chicago, Illinois
- 9-11 American Society of Civil Engineers, Kansas City, Missouri
- 25 American Mathematical Society, New York, New York

November

- 7-9 Conference on Electrical Insulation, Division of Engineering and Industrial Research, National Research Council, Baltimore, Maryland
- 29-30 American Physical Society, Minneapolis, Minnesota

December

- 2-6 American Society of Mechanical Engineers, New York, New York
- 17 Institute of Aeronautical Sciences, New York, New York
- 26-28 Geological Society of America, Chicago, Illinois

Contributed Original Research

A Method for Preparing Rubber Latex Specimens for the Electron Microscope

R. H. KELSEY* AND E. E. HANSON

Chemical and Physical Research Laboratories, The Firestone Tire and Rubber Company, Akron, Ohio

(Received April 26, 1946)

A new method was developed for preparing natural and synthetic rubber specimens for examination with an electron microscope. The technique consisted of mixing latex with a water solution of polyvinyl alcohol and forming a film from the mixture. Electron micrographs of such specimens showed none of the flattening of particles characteristic of the collodion film method and the micrographs were suitable for direct particle-size measurements.

THE preparation of natural and synthetic rubber latex specimens for examination with an electron microscope is difficult because the individual particles are easily deformed and, under the usual conditions of specimen preparation, the particles tend to agglomerate. Von Ardenne and Beischer¹ and others^{2,3} have used a method which consisted of depositing a drop of highly diluted latex on a collodion film and allowing the drop to dry completely. In some instances the drop was withdrawn again into a glass capillary. It was found that in either case the latex particles were flattened at their areas of contact with the collodion film.

The collodion film method was tried in this laboratory with the same poor results observed by Hendricks, Wildman, and McMurdee.² Figure 1 is an electron micrograph of a natural rubber latex particle deposited by evaporating latex diluted 200 to 1 on a collodion film. The film has broken and curled to give a profile view of the particle which is observed to be badly flattened at its area of contact with the film. Figure 2 is a typical electron micrograph of natural rubber

latex particles deposited on a collodion film. The blurred outlines of the particles are explained by the effect shown in Fig. 1. It was obviously impossible to make good particle size measurements from micrographs of this type. It was found further that severe aggregation of particles usually took place when a drop of the



FIG. 1. Natural rubber latex particle. Collodion film broken.

* Present address: Department of Physics, University of Minnesota, Minneapolis, Minnesota.

¹ Von Ardenne and Beischer, "Investigations of the fine structure of high molecular substances with the universal electron microscope," *Rubber Chem. and Tech.* 14, 15 (1941).

² Hendricks, Wildman, and McMurdee, "Morphology of latex particles as shown by electron micrographs," *Ind. Rubber World* 110, 297 (1944).

³ "Electron microscope studies of rubber latices and pigments," *Columbian Carbon Research Laboratories*, April, 1944.

diluted latex was dried down. The latex particles which were left behind when the drop was removed with a glass capillary usually were not badly agglomerated, but it was felt that there was some possibility that the latex particles might adhere selectively according to size to the collodion film. An attempt was made to reduce the amount of agglomeration by incorporating in the diluting water various amounts of such wetting or dispersing agents as sodium oleate, orvis paste, aerosol OT, and ammonium caseinate. None of these materials helped very much in reducing the agglomeration of particles. It was also found in this study that very often the latex particles appeared granular around their edges due to foreign material being swept in during the evaporation of the water and crystallizing out upon the surface of the particle. This effect may also be observed in Fig. 1.

The metallic shadow casting technique used by Williams and Wyckoff⁴ was not tried in this laboratory. However, it is likely that, with this method also, the particles would be flattened at their areas of contact with the glass plate. It is also probable that the particles would have a strong tendency to form aggregates.

In this laboratory, the following approach to the problem was taken: to disperse the latex in a water solution of a film-forming material and to cast a film from it. In this way the latex particles would be completely surrounded by the material of the film and thus the flattening tendency would be greatly reduced. Polyvinyl alcohol, suggested by Mr. R. B. Keller, Jr., proved to be very successful as a water soluble film forming material. From the standpoint of forming strong clear films, the Type B, medium viscosity P.V.A.,* proved to be best. The method consists of the following operations: Into each of two 10-ml beakers is placed 1 ml of 0.5 percent water solution of P.V.A. One drop of the latex is placed in the first beaker and the mixture is stirred vigorously with a glass rod. Approximately two drops of this suspension is

⁴ R. C. Williams and R. W. G. Wyckoff, "The thickness of electron microscopic objects," J. App. Phys. 15, 712 (1944).

⁵ R. C. Williams and R. W. G. Wyckoff, "Applications of metallic shadow casting to microscopy," J. App. Phys. 17, 23 (1946).

* Supplied by E. I. duPont de Nemours and Company.

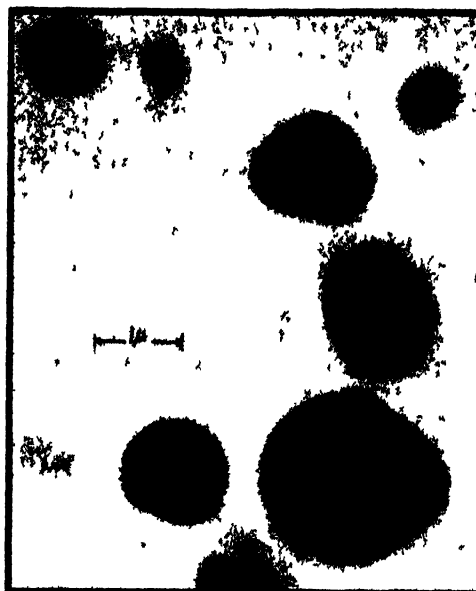


FIG. 2. Natural rubber latex. Collodion film method.

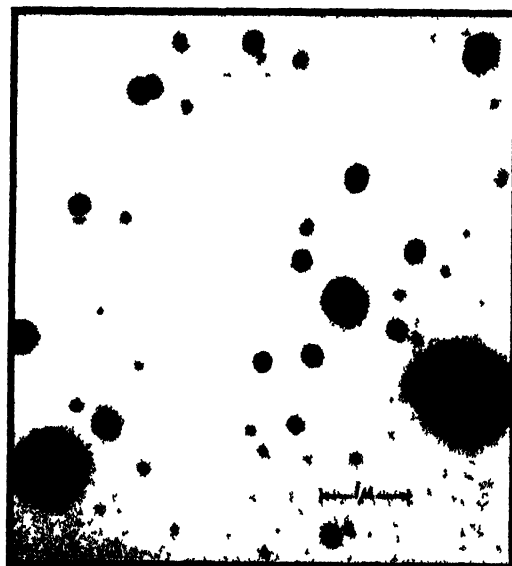


FIG. 3. Natural rubber latex in polyvinyl. Alcohol film.

transferred to the second beaker and the mixture is stirred vigorously. A 200-mesh specimen screen is first dipped in a 0.5 percent solution of aerosol OT, the excess liquid shaken off and the screen then dipped in the latex-P.V.A. suspension. The water component of the film adhering to the specimen screen is allowed to evaporate to dryness and the specimen is placed in the microscope.

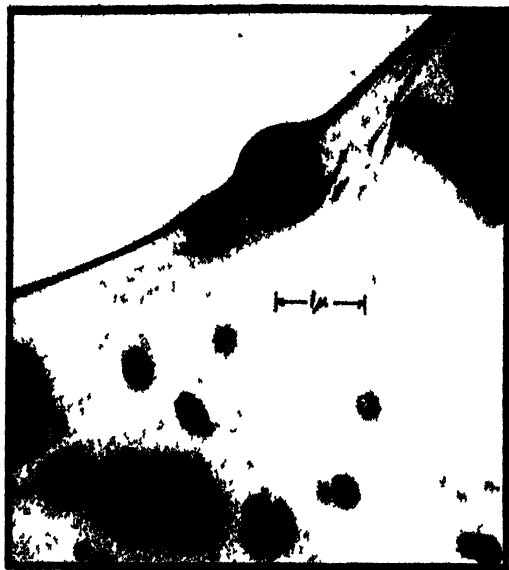


FIG. 4. Natural rubber latex P.V.A. film broken.

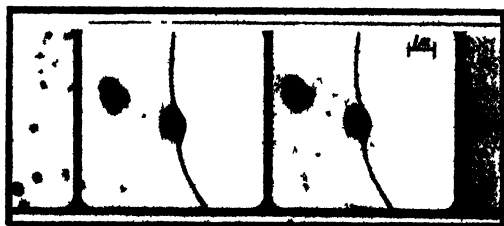


FIG. 5. Stereo electron micrograph of natural rubber latex imbedded in a broken P.V.A. film.

Several "kinks" in the technique are to be noted: The specimen screen was held at its edge with a pair of especially prepared tweezers which could be locked shut with a wire catch. The films of P.V.A. were most quickly dried in a continuously pumped vacuum desiccator with a small air leak for providing air circulation. The use of the wetting agent aerosol proved beneficial in promoting the wetting of the wire mesh with the P.V.A.-latex dispersion. The dilution of the latex had to be varied slightly, of course, depending upon the solids content and the particle size. With a little experience a person can estimate the proper dilution of the latex from the turbidity of the final P.V.A.-latex suspension.

The results obtained by using the P.V.A. technique are indicated in Figs. 3, 4, 5, and 6

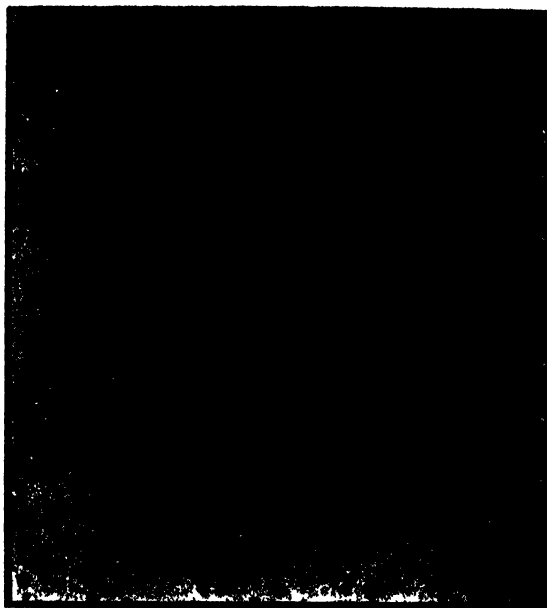


FIG. 6. GR-S type latex in P.V.A. film.

which are micrographs obtained with an RCA Type EMU electron microscope. Figure 3 is a micrograph of natural rubber latex. It is to be noted that the outlines of particles are sharp. Figure 4 is a micrograph of natural rubber latex particles in the edge of a broken P.V.A. film which has curled slightly. Figure 5 is similar to Fig. 4 except that it is a stereo view. There is no evidence of any flattening of the particles. In general, it has been found that whenever the film is stretched, as around a hole in the film, the latex particles tend to elongate in the direction of stretch as is illustrated in Figs. 4 and 5. Figure 6 is a micrograph of a special GR-S type synthetic rubber latex and it illustrates that synthetic rubber latex specimens may also be prepared by the P.V.A. method.

The aggregation of particles is prevented by a violent stirring of the P.V.A.-latex suspension and fewer aggregates occur than with the drying drop method used with collodion film.

It is a pleasure for the authors to acknowledge the interest shown in this problem by Dr. J. H. Dillon and Dr. F. W. Stavely. Thanks are due the Firestone Tire and Rubber Company for releasing this paper for publication.

Perturbation Theory of the Normal Modes for an Exponential M -Curve in Non-Standard Propagation of Microwaves†

C. L. PEKERIS*

Columbia University, New York, New York

(Received May 6, 1946)

In this paper a perturbation method is developed for treating non-standard propagation of microwaves beyond the horizon in the case when the deviation of the M -curve from the standard (=the M -anomaly) can be represented by a term $\alpha e^{-\lambda z}$, where z denotes height in natural units. Here M denotes the modified index of refraction of the air. The method is also applicable to other forms of the M -anomaly which can be derived from an exponential term by differentiation with respect to λ ; in fact, in its region of convergence it is formally applicable to the most general type of M -curve, including elevated ducts. The region of practical convergence of the method ranges from highly substandard conditions down to cases where the decrement is a fraction of the standard value. The procedure followed is to express the height-gain function $U_k(z)$ of the k th mode in the non-standard case as a linear combination of the height-gain functions $U_m^0(z)$ of all the modes in the standard case:

$$U_k(z) = \sum_{m=1}^{\infty} A_{km} U_m^0(z). \quad (a)$$

The execution of this plan hinges on the possibility of evaluating the quantities

$$\beta_{nm}(\lambda) = \int_0^{\infty} U_n^0(z) U_m^0(z) e^{-\lambda z} dz. \quad (b)$$

It is shown that $\beta_{nm}(\lambda)$ satisfies the differential equation

$$\frac{d\beta_{nm}}{d\lambda} = \frac{1}{2\lambda} + \beta_{nm}(\lambda) \cdot \left[-\frac{1}{2\lambda} + \frac{1}{4}(D_m^0 + D_n^0) + \frac{\lambda^2}{4} + \frac{1}{4\lambda^2}(D_m^0 - D_n^0)^2 \right], \quad (c)$$

whose solution is

$$\beta_{nm}(\lambda) = \frac{1}{2\sqrt{\lambda}} \exp \left[\frac{\lambda}{2}(D_n^0 + D_m^0) + \frac{\lambda^3}{12} - \frac{1}{4\lambda}(D_m^0 - D_n^0)^2 \right] \cdot \int_0^{\lambda} \frac{dx}{\sqrt{x}} \exp \left[-\frac{x}{2}(D_m^0 + D_n^0) - \frac{x^3}{12} + \frac{1}{4x}(D_n^0 - D_m^0)^2 \right]. \quad (d)$$

1. INTRODUCTION

IN the theoretical treatment of non-standard propagation of microwaves beyond the horizon by the method of normal modes, one is con-

fronted with the task of solving the equation

$$\beta_{nm} = - \frac{2}{\left[\lambda^2 + 2\lambda(D_m^0 + D_n^0) - 2 + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]} + \frac{8 \left[3\lambda^2 + 2\lambda(D_m^0 + D_n^0) - \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]}{\left[\lambda^2 + 2\lambda(D_m^0 + D_n^0) - 2 + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]^3}. \quad (e)$$

Here D_m^0 denotes the characteristic value of the m th mode in the standard case. For large λ the following asymptotic formula holds

$$\sum_{m=1}^{\infty} A_{km} [(D_k - D_m^0) \delta_{nm} + \alpha \beta_{nm}(\lambda)] = 0, \quad n = 1, 2, 3, \dots, \quad (f)$$

where δ_{nm} denotes the Kronecker symbol. For this purpose a simple iterative procedure has been developed, which has been found to be rapidly convergent. The A_{km} are normalized by the condition

$$\int_0^{\infty} U_k^2(z) dz = 1 = \sum_{m=1}^{\infty} A_{km}^2. \quad (g)$$

[The integral $\int_0^{\infty} U_k^2(z) dz$ diverges when taken along the real axis; it converges, however, and to the same limit, when the path is a radial line in the fourth quadrant of the z plane. In the sequel, whenever an integral is divergent, it will be understood that the path is suitably modified.] One can also expand D_k as a power series in α

$$D_k = D_k^{(0)} + \alpha D_k^{(1)} + \alpha^2 D_k^{(2)} + \dots, \\ D_k^{(1)} = -\beta_{kk}; \quad D_k^{(2)} = e^{i\pi/2} \sum_{m \neq k} \frac{\beta_{mk}^2}{(\tau_m - \tau_k)}, \quad m \neq k.$$

An alternative expression for $D_k^{(2)}$ is given in Eq. (58).

subject to the condition that $U_m(0) = 0$ and that at $h \rightarrow \infty$ U_m should represent an upgoing wave only. Here h denotes height in feet,

$$\frac{d^2 U_m}{dh^2} + k^2 [\gamma(h) + \Lambda_m] U_m = 0, \quad (1)$$

where $N^2(h) - 1 = 2 \times 10^{-6} M(h)$, $k = 2\pi/\lambda$, (2)

where $N(h)$ denotes the modified index of re-

† Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University, New York, New York.

* On leave of absence from the Massachusetts Institute of Technology. This paper is based on work done at Columbia University under contract with the Office of Scientific Research and Development.

fraction of the air and Λ_m is the characteristic value which is generally complex. It is convenient to introduce natural units of height

$$z = h/H, \quad H = (k^2 q)^{-1/2},$$

$$q = \frac{dN^2}{dh} = 2.36 \times 10^{-9} \text{ cm}^{-1}, \quad (3)$$

$$D_m = \Lambda_m (k/q)^{1/2},$$

whereby Eq. (1) is transformed into

$$\frac{d^2 U_m}{dz^2} + [z + f(z) + D_m] U_m = 0. \quad (4)$$

The term $f(z)$ in (4) represents the refraction anomaly, and is equal to zero for a standard atmosphere. In the first instance we shall be treating the case where

$$f(z) = \alpha e^{-\lambda z}, \quad (5)$$

and we shall later generalize the treatment to deal with any M -curve represented as a series of Laguerre functions. If the original M -curve is represented by the expression

$$M(h) = bh + ge^{-ch}, \quad b = .036 \text{ ft.}^{-1}, \quad (6)$$

then α and λ are obtained as follows:

$$\alpha = 2 \times 10^{-6} (k/q)^{1/2} g, \quad \lambda = cH. \quad (7)$$

It is to be noted that in contrast to the constants g and c in Eq. (6) which are independent of frequency, the constants α and λ in Eq. (5) are frequency dependent. For a given observed M -curve the constants α and λ will therefore differ with the frequency band used, as will also the height represented by one unit of z .

2. FORMAL SOLUTION OF THE PROBLEM BY THE PERTURBATION METHOD

In order to solve the equation

$$\frac{d^2 U_k(z)}{dz^2} + [z + \alpha e^{-\lambda z} + D_k] U_k = 0, \quad (8)$$

we seek a solution in the form

$$U_k = \sum_{n=0}^{\infty} A_{kn} U_n^0(z), \quad (9)$$

* No confusion should arise from the use of λ in (5) and the standard usage of λ to denote wave-length.

where $U_n^0(z)$ are the height-gain functions of the n th mode in the standard case, which satisfy the equation

$$d^2 U_n^0(z)/dz^2 + [z + D_n^0] U_n^0(z) = 0, \quad (10)$$

$$\int_0^{\infty} U_n^{2(0)}(z) dz = 1. \quad (11)$$

The solutions of (10) and (11) are well known,

$$U_n^0(z) = C_n u^{1/2} H_{1/2}^{(2)}(u), \quad u = \frac{2}{3}(z + D_n^0)^{3/2}, \quad (12)$$

$$C_n = \left\{ e^{i\pi/4} \frac{2}{\sqrt{3}} \left(\frac{2}{3} \right)^{1/2} v_n^{1/2} [J_{2/3}(v_n) - J_{-2/3}(v_n)] \right\}^{-1}, \quad (13)$$

$$D_n^0 = \tau_n e^{i2\pi/3}, \quad \tau_n = (3v_n/2)^{2/3}, \quad (14)$$

where

$$J_{1/3}(v_n) + J_{-1/3}(v_n) = 0. \quad (15)$$

For small z the power series development of $U_n^0(z)$ is useful:

$$U_n^0(z) = i \sum_{k=1}^{\infty} A_k z^k, \quad (16)$$

$$A_k = -\frac{1}{k(k-1)} [D_n^0 A_{k-2} + A_{k-3}], \quad (17)$$

$$U_n^0(z) = i[z - (D_n^0/6)z^3 - z^4/12 + (D_n^{2(0)}/120)z^5 + (D_n^0/120)z^6 + \dots], \quad (18)$$

while for large z one may use the asymptotic expansion of (16)

$$H_{1/2}^{(2)}(u) \rightarrow \left(\frac{2}{\pi u} \right)^{1/2} e^{i(-u + 5\pi/12)}$$

$$\times \left[1 + \frac{5i}{72u} - \frac{385}{10368u^2} + \dots \right]. \quad (19)$$

If now the expansion (9) be substituted into (8), we obtain, on making use of (10), the condition

$$\sum_{n=1}^{\infty} A_{kn} [(D_k - D_n^0) + \alpha e^{-\lambda z}] U_n^0(z) = 0. \quad (20)$$

On multiplying this equation by $U_n^0(z)$, where n is any integer, and integrating from 0 to ∞ we get a system of equations for the determination of D_k and the A_{kn} :

$$\sum_{n=1}^{\infty} A_{kn} [(D_k - D_n^0) \delta_{nm} + \alpha \beta_{nm}(\lambda)] = 0,$$

$$n = 1, 2, 3, \dots, \quad (21)$$

$$\beta_{nm}(\lambda) = \int_0^\infty U_m^0(z) U_n^0(z) e^{-\lambda z} dz. \quad (22)$$

The characteristic values D_k are then obtained as the roots of the infinite determinant

$$\begin{vmatrix} D_1 - D_1^0 + \alpha\beta_{11}, & \alpha\beta_{12}, & \alpha\beta_{13}, & \dots \\ \alpha\beta_{21}, & D_2 - D_2^0 + \alpha\beta_{22}, & \alpha\beta_{23}, & \dots \\ \alpha\beta_{31}, & \alpha\beta_{32}, & D_3 - D_3^0 + \alpha\beta_{33}, & \dots \\ \dots, & \dots, & \dots, & \dots \end{vmatrix} = 0. \quad (23)$$

Having determined D_k from (23), the A_{km} are obtained by solving the system of linear Eqs. (21).

3. EVALUATION OF $\beta_{nm}(\lambda)$ AS AN INDEFINITE INTEGRAL

The primary task in the perturbation method is the evaluation of the exchange integrals $\beta_{nm}(\lambda)$ defined in Eq. (22). We shall accomplish this by proving that $\beta_{nm}(\lambda)$, as a function of λ , satisfies a differential equation of the first order, for which an explicit solution can be given. For this purpose we shall study the function

$$F(z) \equiv U_n^0(z) U_m^0(z), \quad (24)$$

where

$$\dot{U}_m^0(z) + [z + D_m^0] U_m^0(z) = 0, \quad (25)$$

$$\dot{U}_n^0(z) + [z + D_n^0] U_n^0(z) = 0. \quad (26)$$

By multiplying Eq. (25) by $U_n^0(z)$, Eq. (26) by

$U_m^0(z)$ and subtracting, we obtain

$$\begin{aligned} \frac{d}{dz} (\dot{U}_m^0 U_n^0 - U_m^0 \dot{U}_n^0) \\ = -(D_m^0 - D_n^0) U_m^0 U_n^0, \end{aligned} \quad (27)$$

$$\begin{aligned} \dot{U}_m^0 U_n^0 - U_m^0 \dot{U}_n^0 \\ = -(D_m^0 - D_n^0) \int_0^z U_m^0(x) U_n^0(x) dx. \end{aligned} \quad (28)$$

Now it can be verified by direct substitution that

$$\begin{aligned} \frac{d^2 F}{dz^2} + 2F(D_n + D_m + 2z) + 2F \\ = (D_m^0 - D_n^0) (\dot{U}_m^0 U_n^0 - U_m^0 \dot{U}_n^0) \\ = -(D_m^0 - D_n^0)^2 \int_0^z F(x) dx. \end{aligned} \quad (29)$$

From (29) it follows that

$$F = \frac{d}{dz} \left[(2z + D_m^0 + D_n^0) F + \frac{1}{2} \frac{d^2 F}{dz^2} \right] + \frac{1}{2} (D_m^0 - D_n^0)^2 \int_0^z F(x) dz. \quad (30)$$

We may also note that

$$\frac{d^2 F(0)}{dz^2} = 2 \dot{U}_m^0(0) \dot{U}_n^0(0) = -2, \quad (31)$$

$$\int_0^\infty e^{-\lambda z} \frac{d^2 F}{dz^2} dz = e^{-\lambda z} (F + \lambda F) \Big|_0^\infty + \lambda^2 \int_0^\infty e^{-\lambda z} F dz = \lambda^2 \int_0^\infty e^{-\lambda z} F dz, \quad (32)$$

$$\int_0^\infty e^{-\lambda z} dz \int_0^z F(x) dx = -\frac{1}{\lambda} e^{-\lambda z} \int_0^z F(x) dx \Big|_0^\infty + \frac{1}{\lambda} \int_0^\infty e^{-\lambda z} F(z) dz = \frac{1}{\lambda} \int_0^\infty e^{-\lambda z} F(z) dz. \quad (33)$$

$$\int_\lambda^\infty e^{-\lambda z} F(z) dz = -\frac{d}{d\lambda} \int_\lambda^\infty e^{-\lambda z} F(z) dz. \quad (34)$$

* This is the first occasion in the author's experience where use is made of the fact that the product of two functions, each of which is a solution of a distinct second-order ordinary linear differential equation, satisfies a fourth-order linear differential equation. See G. N. Watson, *Theory of Bessel Functions* (The Macmillan Company, New York, 1944), p. 146.

We now substitute Eq. (30) in the integrand of (22) and obtain

$$\begin{aligned}
 \beta_{nm}(\lambda) &= \int_0^\infty e^{-\lambda z} F(z) dz = \int_0^\infty e^{-\lambda z} dz \left\{ \frac{d}{dz} \left[(2z + D_m^0 + D_n^0) F + \frac{1}{2} \frac{d^2 F}{dz^2} \right] + \frac{1}{2} (D_m^0 - D_n^0)^2 \int_0^z F(x) dx \right\} \\
 &= \frac{1}{2\lambda} (D_m^0 - D_n^0)^2 \int_0^\infty e^{-\lambda z} F(z) dz + \left[(2z + D_m^0 + D_n^0) F + \frac{1}{2} \frac{d^2 F}{dz^2} \right] e^{-\lambda z} \Big|_0^\infty \\
 &\quad + \lambda \int_0^\infty e^{-\lambda z} \left[(2z + D_m^0 + D_n^0) F + \frac{1}{2} \frac{d^2 F}{dz^2} \right] dz \\
 &= 1 + \int_0^\infty e^{-\lambda z} F(z) \left[2\lambda z + \lambda (D_m^0 + D_n^0) + \frac{1}{2} \lambda^2 + \frac{1}{2\lambda} (D_m^0 - D_n^0)^2 \right] dz \\
 &= 1 - 2\lambda \frac{d\beta_{nm}(\lambda)}{d\lambda} + \beta_{nm}(\lambda) \left[\lambda (D_m^0 + D_n^0) + \frac{\lambda^2}{2} + \frac{1}{2\lambda} (D_m^0 - D_n^0)^2 \right]. \quad (35)
 \end{aligned}$$

It follows that the exchange integral $\beta_{nm}(\lambda)$ satisfies the first-order differential equation

$$\frac{d\beta_{nm}(\lambda)}{d\lambda} = \frac{1}{2\lambda} + \beta_{nm}(\lambda) \cdot \left[-\frac{1}{2\lambda} + \frac{1}{2} (D_m^0 + D_n^0) + \frac{\lambda^2}{4} + \frac{1}{4\lambda^2} (D_m^0 - D_n^0)^2 \right]. \quad (36)$$

The solution of (36) is

$$\begin{aligned}
 \beta_{nm}(\lambda) &= \frac{1}{2\sqrt{\lambda}} \exp \left[\frac{\lambda}{2} (D_m^0 + D_n^0) + \frac{\lambda^3}{12} - \frac{1}{4\lambda} (D_m^0 - D_n^0)^2 \right] \\
 &\quad \cdot \int_0^\lambda \frac{dx}{\sqrt{x}} \exp \left[-\frac{x}{2} (D_m^0 + D_n^0) - \frac{x^3}{12} + \frac{1}{4x} (D_m^0 - D_n^0)^2 \right]. \quad (37)
 \end{aligned}$$

4. PROPERTIES OF $\beta_{nm}(\lambda)$

For small λ the solution of the differential Eq. (36) can be started with a power series in λ .

(a). $n \neq m$

$$\beta_{nm}(\lambda) = -\frac{2\lambda e^{i2\pi/3}}{(\tau_m - \tau_n)^2} \sum_{k=0}^{\infty} C_k e^{i2\pi k/3} \lambda^k, \quad (38)$$

$$\begin{aligned}
 C_0 &= 1, \quad C_1 = 6/(\tau_m - \tau_n)^2, \\
 C_2 &= [10C_1 - 2(\tau_m + \tau_n)]/(\tau_m - \tau_n)^2, \\
 C_3 &= [14C_2 - 2C_1(\tau_m + \tau_n)]/(\tau_m - \tau_n)^2, \\
 C_n &= [(4n+2)C_{n-1} - 2(\tau_m + \tau_n)C_{n-2} - C_{n-4}]/(\tau_m - \tau_n)^2.
 \end{aligned} \quad (39)$$

(b). $n = m$

$$\beta_{mm} = 1 + B_1\lambda + B_2\lambda^2 + \dots, \quad (40)$$

$$B_1 = \frac{4}{15} D_m^0, \quad B_2 = \frac{4}{15} D_m^2(0), \quad B_3 = \frac{1}{14} + \frac{8}{105} D_m^3(0), \quad B_n = \frac{1}{(2n+1)} [2D_m^0 B_{n-1} + \frac{1}{2} B_{n-3}]. \quad (41)$$

† When this paper was first reported at a conference in June, 1945, Dr. W. F. Eberlein commented that he had independently derived Eq. (36) by an application of the theory of the Laplace transform.

For intermediate values of λ one may either use the integral in (37) or one may integrate numerically the differential Eq. (36). The latter procedure was advocated by Professor Hartree.

For large values of λ an asymptotic expansion can be obtained directly from (36) by writing it in the form

$$\beta_{nm}(\lambda) = \frac{-2 + 4\lambda \frac{d\beta_{nm}(\lambda)}{d\lambda}}{\left[\lambda^3 + 2\lambda(D_m^0 + D_n^0) - 2 + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]} - \frac{2}{\left[\lambda^3 + 2\lambda(D_m^0 + D_n^0) - 2 + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]} + \frac{8 \left[3\lambda^3 + 2\lambda(D_m^0 + D_n^0) - \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]}{\left[\lambda^3 + 2\lambda(D_m^0 + D_n^0) - 2 + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]^3}. \quad (42)$$

An alternative asymptotic expansion can be derived from (37) by partial integration

$$\beta_{mn} \rightarrow -\frac{2}{\left[\lambda^3 + 2\lambda(D_m^0 + D_n^0) + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]} + \frac{4 \left[5\lambda^3 + 2\lambda(D_m^0 + D_n^0) - \frac{3}{\lambda}(D_m^0 - D_n^0)^2 \right]}{\left[\lambda^3 + 2\lambda(D_m^0 + D_n^0) + \frac{1}{\lambda}(D_m^0 - D_n^0)^2 \right]^3}. \quad (43)$$

In doing so one needs to prove that

$$\int_0^\infty \frac{dx}{\sqrt{x}} \exp \left[-\frac{x}{2}(D_m^0 + D_n^0) - \frac{x^3}{12} + \frac{1}{4x}(D_m^0 - D_n^0)^2 \right] \equiv \psi_{nm} = 0. \quad (44)$$

We shall state here without proof that

$$\psi_{mm} = \int_0^\infty \frac{dx}{\sqrt{x}} \exp \left[-xD_m^0 - \frac{x^3}{12} \right] = \frac{\pi\sqrt{\pi}}{2} \left(\frac{2}{3} \right)^{1/3} h_1(D_m^0) h_2(D_m^0), \quad (45)$$

where h_1 and h_2 are Furry's functions of the first and second kind defined as

$$h_1(x) = \left(\frac{2}{3} \right)^{1/3} x^{1/3} H_{1/3}^{(1)} \left(\left(\frac{2}{3} \right)^{1/3} x^{1/3} \right), \quad (46)$$

$$h_2(x) = \left(\frac{2}{3} \right)^{1/3} x^{1/3} H_{1/3}^{(2)} \left(\left(\frac{2}{3} \right)^{1/3} x^{1/3} \right). \quad (47)$$

Since by definition of D_m^0 , $h_2(D_m^0) = 0$, it follows that $\psi_{mm} = 0$. The proof of (44) for $n \neq m$ is left as an exercise to the interested reader.

5. ITERATION METHOD OF SOLVING FOR THE CHARACTERISTIC VALUES D_k AND THE COEFFICIENTS A_{km}

In solving Eqs. (21) and (23), which are of infinite order, one proceeds by first assuming that $A_{km} = 0$ for $m > p$, where p is a convenient integer, and then evaluating D_k and A_{km} , $m = 1, 2, \dots, p$. Next, one assumes that $A_{km} = 0$ for $m > p+1$ resolves for D_k and the A_{km} , and the accuracy of the results is judged by the agreement between the values in successive approximations. The direct solution of (23) and (21) is, however, a laborious process which rapidly increases in complexity as p exceeds about 4. The following iterative procedure has been found effective and of the same intrinsic simplicity for any value of p .

To begin with, the p equations in (21), being homogeneous, do not determine the absolute values of all the A_{km} but merely the ratios of $(p-1)$ of them to a p th one. The absolute values are then determined from the normalization condition

$$\int_0^\infty U_k^2(z) dz = 1 = \sum_{m=1}^\infty A_{km}^2. \quad (g)$$

Let therefore

$$C_{km} = A_{km}/A_{kk}, \quad C_{kk} = 1, \quad (48)$$

and the p equations in (21) are just sufficient to determine the $(p-1)$ constants C_{km} and D_k . We divide the equations in (21) by A_{kk} and pick the k th equation ($n=k$) to solve for D_k , while the other equations are used to solve for the C_{km} , as is illustrated in the scheme below for the particular case of $k=1$.

$$\frac{D_1}{\alpha} = \frac{D_1^0}{\alpha} - \beta_{11} - C_{12}\beta_{12} - C_{13}\beta_{13} - \dots, \quad (49)$$

$$\left(\frac{D_1}{\alpha} - \frac{D_2^0}{\alpha} + \beta_{22} \right) C_{12} = -\beta_{12} - C_{13}\beta_{23} - C_{14}\beta_{24} - \dots, \quad (50)$$

$$\left(\frac{D_1}{\alpha} - \frac{D_3^0}{\alpha} + \beta_{33} \right) C_{13} = -\beta_{13} - C_{12}\beta_{23} - C_{14}\beta_{34} - \dots, \quad (51)$$

$$\left(\frac{D_1}{\alpha} - \frac{D_4^0}{\alpha} + \beta_{44} \right) C_{14} = -\beta_{14} - C_{12}\beta_{24} - C_{13}\beta_{34} - \dots. \quad (52)$$

As a first approximation one puts

$$\frac{D_1}{\alpha} = \frac{D_1^0}{\alpha} - \beta_{11}, \quad (53)$$

$$C_{12} = -\beta_{12} / \left(\frac{D_1}{\alpha} - \frac{D_2^0}{\alpha} + \beta_{22} \right), \quad (54)$$

$$C_{13} = -\beta_{13} / \left(\frac{D_1}{\alpha} - \frac{D_3^0}{\alpha} + \beta_{33} \right), \text{ etc.}, \quad (55)$$

where the value of D_1/α obtained from (53) is used in (54) and (55). Next, one substitutes these values of the C 's in the right-hand sides of Eqs. (49) to (52) and resolves for D_1/α and the C 's. This procedure has been found to be rapidly convergent, and is, furthermore, self-correcting in case of arithmetical errors.

6. EXPANSION OF D_k INTO A POWER SERIES IN α

When α is small, it is convenient to expand D_k into the series

$$D_k = D_k^{(0)} + \alpha D_k^{(1)} + \alpha^2 D_k^{(2)} + \dots. \quad (56)$$

It is known from standard perturbation theory that

$$D_k^{(1)} = -\beta_{kk}; \quad D_k^{(2)} = e^{i\pi/3} \sum_m \frac{\beta_{mk}^2}{(\tau_m - \tau_k)}, \quad m \neq k. \quad (57)$$

It is possible also to derive an alternative expression for $D_k^{(2)}$:

$$\begin{aligned}
 D_k^{(2)}(\lambda) &= \frac{1}{2} \lambda D_k^{(1)}(\lambda)^2 + \frac{1}{2\sqrt{\lambda}} \exp \left[D_k^{(0)} \lambda + \frac{\lambda^2}{12} \right] \int_0^\lambda \exp \left[-D_k^{(0)} x - \frac{x^2}{12} \right] \\
 &\quad \cdot \left[\left(1 + \frac{\lambda}{2x} \right) D_k^{(1)}(\lambda+x) + D_k^{(1)}(\lambda) D_k^{(1)}(x) \right] x^{\frac{1}{2}} dx \\
 &= \lambda D_k^{(1)}(\lambda)^2 + \frac{1}{4\sqrt{\lambda}} \exp \left[D_k^{(0)} \lambda + \frac{\lambda^2}{12} \right] \int_0^\lambda \left[D_k^{(1)}(\lambda) + \left(2 + \frac{\lambda}{x} \right) D_k^{(1)}(\lambda+x) \right] \\
 &\quad \cdot \exp \left[-D_k^{(0)} x - \frac{x^2}{12} \right] x^{\frac{1}{2}} dx. \quad (58)
 \end{aligned}$$

Since the former expression is simpler for computational purposes we shall not give here the derivation of (58).

7. APPLICABILITY OF THE PERTURBATION METHOD TO A MORE GENERAL CLASS OF M -ANOMALIES

It is possible to apply the results obtained for the case when the M -anomaly is of the form $f(z) = \alpha e^{-\lambda z}$ to more general types of M -anomalies. To begin with, if

$$f(z) = \alpha e^{-\lambda z} + \gamma e^{-\mu z}, \quad (59)$$

then we merely write in Eq. (21) in place of $\alpha \beta_{nm}(\lambda)$, $[\alpha \beta_{nm}(\lambda) + \gamma \beta_{nm}(\mu)]$. Once the $\beta_{nm}(\lambda)$ are computed as functions of λ , there is no additional labor required to deal with an $f(z)$ which consists of a sum of any number of exponential terms. If instead of $f(z) = \alpha e^{-\lambda z}$ we had $f(z) = \alpha z e^{-\lambda z}$, then the corresponding $\beta_{nm}'(\lambda)$ would be

$$\begin{aligned}
 \beta_{nm}'(\lambda) &= \int_0^\infty U_m^0(z) U_n^0(z) z e^{-\lambda z} dz \\
 &= - \frac{d\beta_{nm}(\lambda)}{d\lambda}. \quad (60)
 \end{aligned}$$

If $\beta_{nm}(\lambda)$ is known, $d\beta_{nm}(\lambda)/d\lambda$ can be computed directly from Eq. (36). When Eq. (36) is integrated numerically, the derivative $d\beta_{nm}(\lambda)/d\lambda$ is computed at each point in any case. Evidently,

for $f(z) = \alpha z^k e^{-\lambda z}$, where k is a positive integer,

$$\begin{aligned}
 \beta_{nm}'(\lambda) &= \int_0^\infty U_m^0(z) U_n^0(z) z^k e^{-\lambda z} dz \\
 &= (-)^k \frac{d^k \beta_{nm}(\lambda)}{d\lambda^k}. \quad (61)
 \end{aligned}$$

By successive differentiation of Eq. (36), it is possible to express any high order derivative of $\beta_{nm}(\lambda)$ in terms of $\beta_{nm}(\lambda)$. From a purely formal point of view we can say therefore that by our method we can treat any M -anomaly by expanding it into a series of Laguerre functions, since these functions involve only terms of the form $z^k e^{-\lambda z}$. It may be pointed out that a single term $z^k e^{-\lambda z}$ vanishes both at the ground and at great height and reaches a maximum at $z = k/\lambda$. Such a single term is therefore suitable to represent an elevated duct.

Tables of $\beta_{nm}(\lambda)$ for n and m ranging from 1 to 6 are available at the Columbia University Division of War Research, Mathematical Physics Group, for $0 < \lambda(0.1) < 4$ and $0 < 1/\lambda(0.1) < .25$. These tables were computed by Mrs. B. Brown, and Misses D. Weinstock, L. Selig, F. Jones, and E. Herman under the supervision of Miss Alice Osterberg.

Stresses in Cylindrical Glass-Metal Seals with Glass Inside

ALBERT W. HULL

Research Laboratory, General Electric Company, Schenectady, New York

(Received March 28, 1946)

It was shown previously that, when a cylinder of glass is sealed to the outside of a metal rod, the principal stresses in the glass are of opposite sign, so that tensile stresses cannot be avoided except by a perfect match. In this article the stresses are calculated for a solid glass cylinder sealed to the *inside* of a metal cylinder. It is shown that the stresses are *all of the same sign*, so that a moderate mismatch in thermal expansion, with the metal expansion the greater, is allowable and perhaps desirable. Large differences in expansion should be avoided, because of the shearing stresses at the ends.

MANY modern devices use cylindrical seals, in which a cylinder of glass is sealed to the inside surface of a metal cylinder, either with or without a central wire. These devices include microwave tubes, thyratrons, capacitors, and refrigerators. The stresses in such seals are different from those in glass-wire seals. They can be calculated easily from the equations given in an earlier paper.¹

CALCULATION OF STRESSES

The form of seal is shown in Fig. 1. Let the components of stress be p_r , p_θ , and p_z , in the radial, tangential, and axial directions, respectively. Only the central portion of the seal will be considered, in which the stresses are independent of z , and in which no shearing stresses exist. The equations of the earlier paper, which refer to stresses in the outer cylinder, are:

Radial stress,

$$p_r = - \left[\frac{E_2 \delta}{1 + \alpha + \alpha \beta R} \right] \left[\frac{a^2}{b^2} - \frac{a^2}{r^2} \right], \quad (1)$$

Tangential stress,

$$p_\theta = - \left[\frac{E_2 \delta}{1 + \alpha + \alpha \beta R} \right] \left[\frac{a^2}{b^2} + \frac{a^2}{r^2} \right], \quad (2)$$

Axial stress,

$$p_z = - \left[\frac{E_2 \delta}{1 + \alpha + \alpha \beta R} \right] \left[2\sigma \frac{a^2}{b^2} + \frac{1 + \alpha + \alpha \beta R}{1 + \beta R} \right], \quad (3)$$

where

E_1 = elasticity (Young's modulus) of internal cylinder,

¹ A. W. Hull and E. E. Burger, "Glass to metal seals," *Physics* 5, 387 (1934).

E_2 = elasticity (Young's modulus) of external cylinder, σ = Poisson's ratio, assumed equal for metal and glass, a , b = radii of internal and external cylinders, respectively,

$$R = E_2/E_1, \quad \alpha = (a^2/b^2)(1 - 2\sigma), \quad \beta = b^2/a^2 - 1,$$

$$\delta = (k_2 - k_1)(t - t_0),$$

k_2 and k_1 are the mean contraction coefficients between t_0 and t of the outer cylinder and inner cylinder, respectively,

t is the temperature at which the stress is observed, and t_0 is the "sealing temperature," i.e., the temperature at which the glass ceases to flow after sealing. δ is, therefore, the difference in total contraction of metal and glass, in cooling from the sealing temperature to the observation temperature (normally room temperature). The stress is defined to be positive when it is tension, negative when compression.

These equations may be modified to represent the stresses of the inner cylinder by making use of three simple relations²:

(a) The radial stress is the same for inner and outer cylinders at the boundary $r = a$, and is constant throughout the inner cylinder.

(b) The tangential and radial stresses are equal in the inner cylinder.

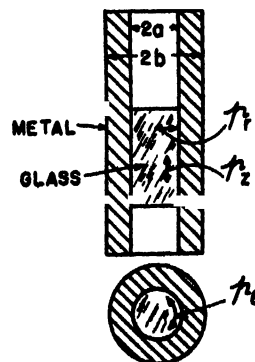


FIG. 1. Cylindrical seal with glass inside. The stress in the central portion can be calculated from expansion data.

² H. Poritsky, "Analysis of thermal stresses in sealed cylinders," *Physics* 5, 406 (1934).

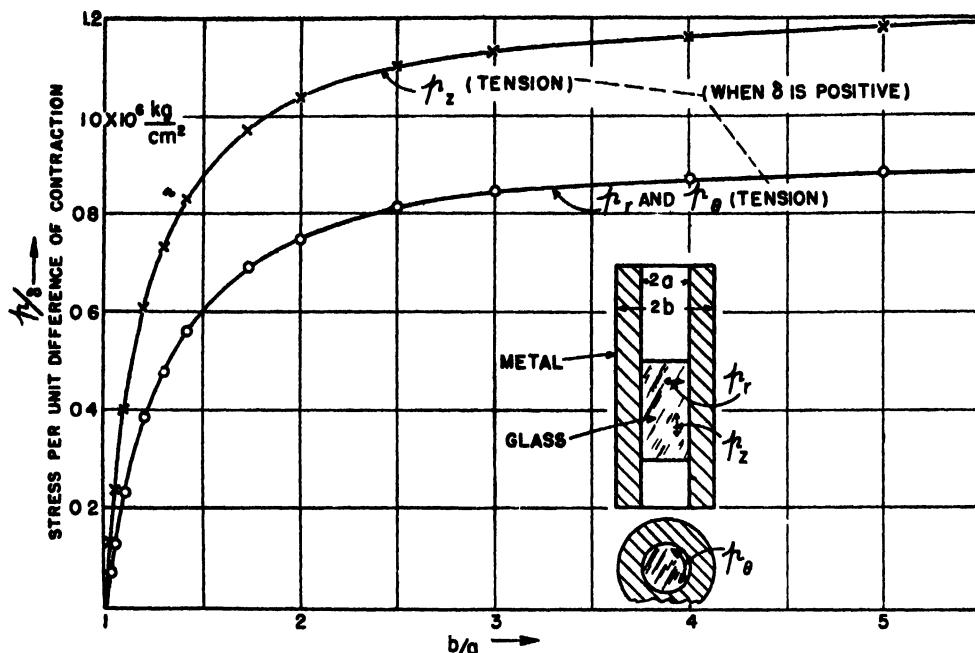


FIG. 2. Stress in internal glass cylinder per unit difference of contraction, δ , between metal and glass ($\delta = \{k_{\text{outer cyl}} - k_{\text{inner cyl}}\} (t - t_0)$), as function of ratio b/a of metal diameter to glass diameter. When δ is negative, the stresses will be compression.

(c) The total axial stress over any cross section $z = \text{constant}$ of both cylinders is zero. Hence,

$$p_r(\text{internal}) = -\frac{b^2 - a^2}{a^2} p_r(\text{external}).$$

The calculations will be made for an external cylinder of iron, for which $E_2 = 2.05 \times 10^6$ kg/cm². This is the metal most likely to be used in thick sections, i.e., with large values of b/a . For Fernico

TABLE I. Stresses in internal glass cylinder per unit difference of contraction δ , between metal and glass, as function of ratio b/a of metal diameter to glass diameter. Positive stresses are tension; negative, compression. When δ is negative, the signs of all stresses are opposite to those given in the table.

b/a	$P_r = P_\theta$	P_z
∞	+0.91	+1.193
5	+0.882	+1.18
4	+0.872	+1.164
3	+0.845	+1.130
2.5	+ .811	+1.102
2.0	+ .750	+1.040
1.732	+ .691	+0.975
1.414	+ .560	+0.830
1.30	+ .460	+ .735
1.20	+ .388	+ .610
1.10	+ .230	+ .401
1.05	+ .1285	+ .236
1.025	+ .071	+ .1325
1.00	0.000	0.000

($E_2 = 1.80 \times 10^6$), with thicknesses ordinarily used ($b/a \leq 1.1$), the radial and tangential stresses are about 2 percent greater than those given here, and the axial stress 4 percent greater; while the maximum difference, even for very thick Fernico tubing, does not exceed 10 percent. The elastic constants used for the calculations are therefore:

$$\begin{aligned} E_1 (\text{glass}) &= 0.65 \times 10^6 \text{ kg/cm}^2, \\ E_2 (\text{metal}) &= 2.05 \times 10^6 \text{ kg/cm}^2, \\ \sigma_1 = \sigma_2 &= 0.30. \end{aligned}$$

The calculated values for the stresses in the inner cylinder are given in Table I and Fig. 2. For comparison, the stresses in external glass cylinders are shown in Fig. 3.¹

COMPARISON OF INTERNAL AND EXTERNAL SEALS

A comparison of Figs. 2 and 3 reveals one striking difference. The stresses in internal seals are all of the same sign. Thus, when the contraction of the metal is greater than that of the glass (δ negative) all the stresses will be negative, i.e., compressions. This is a desirable condition, because glass is strong in compression. However, excessive differences of contraction should be

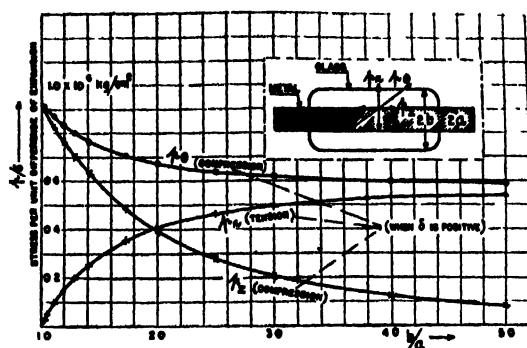


FIG. 3. Stress in cylindrical seal at surface of wire ($r=a$) per unit difference of contraction, δ , between metal and glass, ($\delta = \{k_{\text{outer cyl}} - k_{\text{inner cyl}}\} \{t - t_0\}$) as function of ratio b/a of glass diameter to wire diameter. When δ is negative, the signs of stresses are the reverse of those indicated, viz. tension in place of compression, and vice versa.

avoided, since the shearing stresses at the ends are proportional to the difference in contraction.

With external seals the situation is different. Here the two largest stresses, namely, the radial and the tangential stresses, are of opposite sign, so that there is no safe condition except that of zero stress. For example, a metal which contracts more than the glass will give an axial stress which is a compression, as observed with the

quartz wedge testing equipment²; and this is sometimes mistaken for a desirable mismatch. Actually, as may be seen in Fig. 3, the tangential stress also is a compression, but the radial stress is tension, tending to break the bond between metal and glass. This bond is the weakest part of the seal; hence this type of stress should be avoided.

INTERNAL SEAL WITH CENTRAL WIRE

When a central wire is added to the internal seal (Fig. 4), one has a combination of the two cases considered above, namely an internal and an external seal. For the central wire seal, in which the glass is external, the only safe condition is a close match to the glass. The external metal cylinder, however, may be mismatched to yield a mild compression of the glass. The values given in Table I and Fig. 2 will apply to this case, with small error, provided the central wire is well matched to the glass.



FIG. 4. Cylindrical seal with central wire.

² A. W. Hull and E. E. Burger, *Rev. Sci. Inst.* 7, 98 (1936).

Particle Size Determination from X-Ray Line Broadening

L. S. BIRKS AND H. FRIEDMAN
U. S. Naval Research Laboratory, Washington, D. C.
(Received April 8, 1946)

The x-ray line broadening method of determining particle size was compared with direct measurement on electron micrographs. By controlled heating of the carbonate, magnesium oxide particles were prepared from 50 to 1000 Å in diameter. Particle size calculated from x-ray data taken on a Geiger counter spectrometer agreed to ± 10 percent with the microscope measurements. Mechanical mixtures of two different sizes were examined by the x-ray method, but the particle sizes could not be determined unless the two maxima of the distribution curve were completely resolved.

INTRODUCTION

THE broadening of x-ray diffraction lines is one of the most accurate indirect methods of determining particle size for crystallites smaller than 1000 Å diameter. The theory relating line broadening to particle size was developed by Scherrer,¹ von Laue,² and Bragg.³ It was con-

sidered in more detail by Warren,⁴ Patterson,⁵

Nach. Gesell. Wiss. Göttingen, *Sitzungsber.*, July 26, 1918 in R. Zsigmondy, *Kolloidchemie* (Otto Spamer, Leipzig, 1920), third edition.

² Von Laue, *Zeits. f. Krist.* 64, 115 (1926).

³ W. Bragg, *The Crystalline State* (G. Bell and Sons, London, 1933).

⁴ B. E. Warren, "X-ray diffraction study of carbon blacks," *J. Chem. Phys.* 2, 551 (1934).

⁵ A. L. Patterson, "The diffraction of x-rays by small crystalline particles," *Phys. Rev.* 56, 972 (1939).

¹ Scherrer, "Bestimmung der Grosse und der inneren Struktur von Kolloidteilchen mittels Röntgenstrahlen,"

Jones,⁶ Murdock,⁷ and others. Warren⁸ in particular showed that the size of MgO particles could be determined from x-ray line broadening measurements. In the present paper, the particle size of MgO determined from x-ray line broadening was compared with the size by direct measurement with the electron microscope over the range from 50 to 1000Å. Magnesium oxide was chosen for the particle size studies because of the ease with which uniform specimens could be prepared by heating the carbonate.

DIRECT MEASUREMENT WITH THE ELECTRON MICROSCOPE

Magnesium oxide was prepared by heating magnesium carbonate in air. The temperature of conversion was varied from 200°C to 1000°C and heating periods ranged from 2 hours to 12 hours. The electron microscope was calibrated with a grating replica. The replica spacing was determined to ± 5 percent using the lines of the mercury spectrum. Nineteen different fields were photographed in the electron microscope and the mean deviation was 1 percent. Each of the oxide specimens was then photographed at the same

magnification as the grating replica. Therefore the accuracy of the measured magnification was estimated to be ± 5 percent.

Figures 1 through 6 are typical micrographs of the unconverted carbonate and the oxide at various temperatures. Originally the carbonate was in the form of thin flakes of the order of

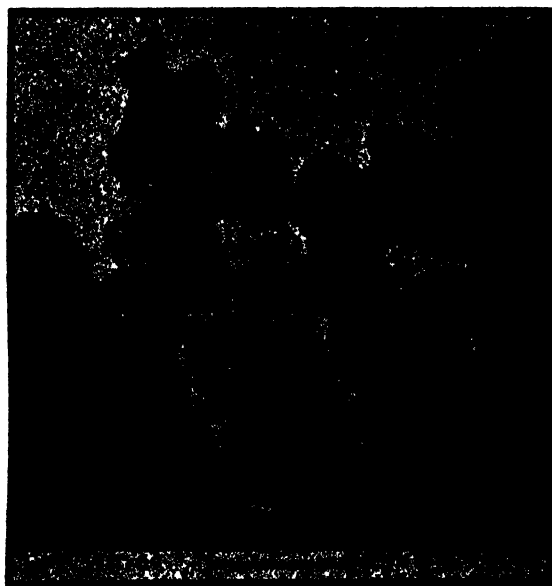


FIG. 1. Commercial magnesium carbonate. Scale indicates one micron.

⁶ F. W. Jones, Proc. Roy. Soc. A166, 16 (1938).

⁷ C. C. Murdock, Phys. Rev. 36, 8 (1930).

⁸ B. E. Warren, J. App. Phys. 12, 375 (May 1941).

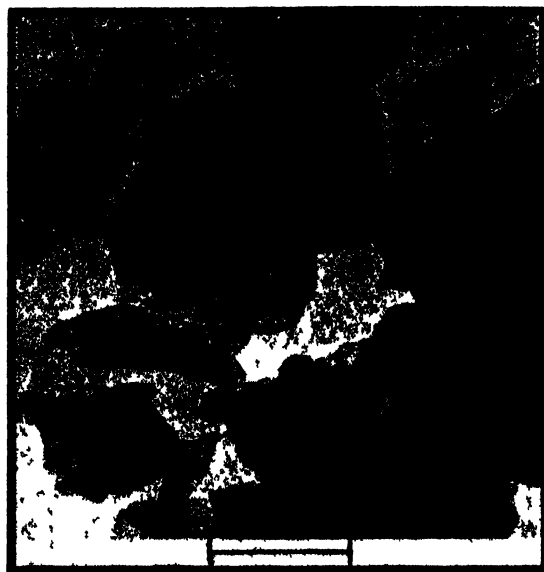


FIG. 2. Magnesium carbonate heated at 200°C for 6 hours. Scale indicates one micron.

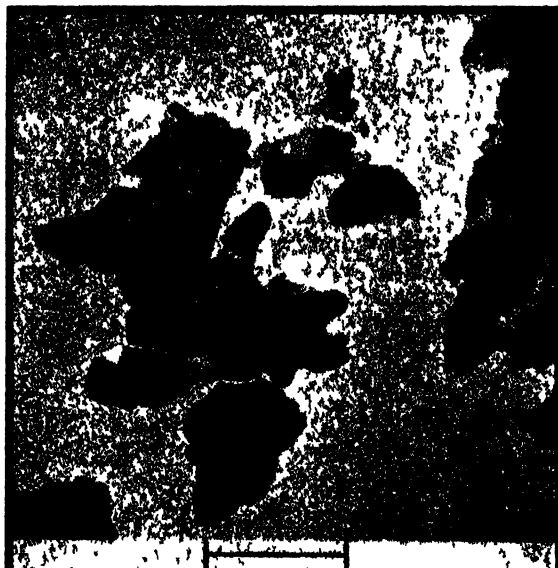


FIG. 3. Magnesium carbonate converted to magnesium oxide at 400°C for 6 hours. Scale indicates one micron.

one micron in diameter. After treatment at 200°C no apparent change had taken place. At 400°C however, a definite mosaic appeared in the flakes and it was possible to make a rough estimate of the size of the particles comprising the mosaic. The accuracy of the measurement was not better than ± 25 percent. At high temperature the particles were larger and clearly separated although the shape of the original carbonate flakes was still apparent. For each heat treatment, several hundred particles were measured and distribution curves plotted. Each specimen showed a narrow distribution curve with only slight asymmetry. The peak value in

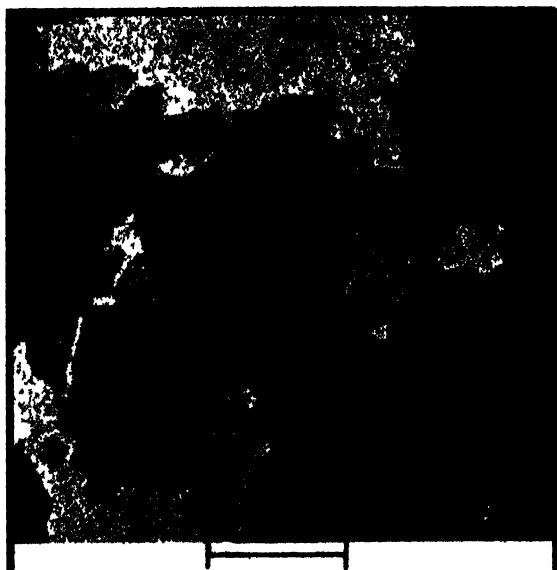


FIG. 4. Magnesium carbonate converted to magnesium oxide at 600°C for 6 hours. Scale indicates one micron.

angstroms plotted against heating period for the various temperatures is shown by the dotted lines in Fig. 7. The sizes range from 50Å at 400°C and 2 hours to 1000Å at 1000°C and 12 hours. The rate of growth was slow at 400°C but increased rapidly with increasing temperature to a value of approximately 40Å per hour at 1000°C. These results are compared below with those obtained by x-ray diffraction.

X-RAY LINE BROADENING

The determination of size by x-ray diffraction makes use of the relation that for particles

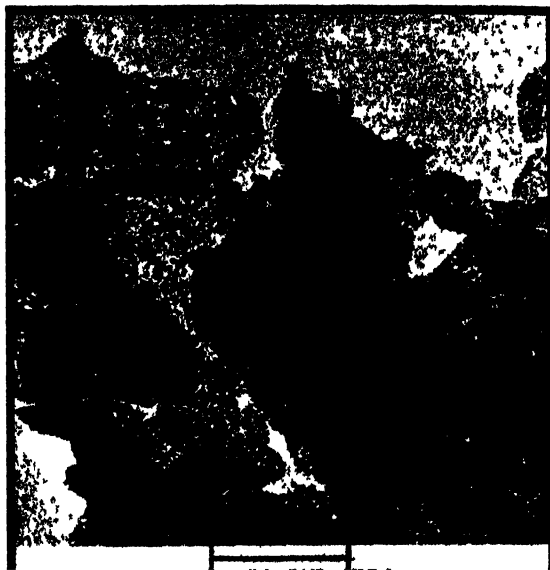


FIG. 5. Magnesium carbonate converted to magnesium oxide at 800°C for 6 hours. Scale indicates one micron.



FIG. 6. Magnesium carbonate converted to magnesium oxide at 1000°C for 6 hours. Scale indicates one micron.

smaller than about 1000Å the diffraction lines are broadened by an amount depending on the size of the particles. The width of the diffraction lines from large particles are a characteristic of the apparatus and do not depend on the particle size. An analogous effect is found in optical grating spectra; as the number of lines per inch

is increased, the diffraction pattern of visible light becomes sharper and sharper until the resolving power of the measuring apparatus is reached. In x-ray diffraction, planes of atoms correspond to the lines in an optical grating. For the type of x-ray diffraction equipment usually employed, a specimen with particles of the order of 1000A or larger gives the sharpest possible maxima.

The problem of relating the broadening of the line to the size of the particles has been treated by many investigators.¹⁻⁸ In general, their methods all lead to an equation of the form,

$$D = K\lambda/B \cos \theta, \quad (1)$$

where

D = diameter of the particle,
 K = constant (value given below),
 B is the broadening of the line,
 λ is the wave-length of the x-radiation, and
 θ is the Bragg angle.

The broadening B may be determined in several ways. The method of Scherrer¹ was to measure the width of the diffraction line in radians at a point where the intensity was half the maximum value. Von Laue² measured the area under the

diffraction line and divided by the maximum intensity. This is known as the integral width method. Recent investigators^{5,6} adopted more rigorous derivations which took into account the shape of the distribution curves for the particles and the direction of measurement in the crystal. The diffraction lines from the MgO specimens in this investigation were symmetrical and their shape closely approximated an error function. The method of width at half maximum was used for all the determinations of broadening. No matter which equation is adopted, B must be corrected for the width of the diffraction line for large particles, known as the instrument width. The correction takes the form,

$$B^2 = B'^2 - b'^2, \quad (2)$$

where B is again the broadening, B' is the width of the line from the small particles, and b' is the instrument width or the width of the line from particles larger than 1000A. The constant K in the equation depends on the method of measuring B and on the assumptions as to the shape of the particle. Scherrer¹ arrived at a value of $K=0.89$. Patterson⁵ obtained a value of $K=1.11$ when the particles were assumed to be spherical in shape and of cubic structure. From the electron micrographs, the particles treated at 1000°C appeared to be neither strictly spherical nor cubic, although the structure of MgO is cubic. A value of $K=1$ was adopted as a compromise.

The x-ray specimens were prepared by mixing approximately $\frac{1}{2}$ gram of the material with a 2 percent solution of parlodion in amyl acetate. The mixture was smeared on a microscope slide to give a flat surface approximately 1 inch square. The N.R.L. Geiger counter spectrometer⁹ was employed with a divergent beam and focusing geometry to give high intensity. The Geiger counter pulses were fed into either a scaling circuit for measuring the intensity at discreet points along the pattern or into a recorder for automatic scanning of the complete pattern.

Each specimen was first scanned and the recorded patterns are shown in Fig. 8. The carbonate pattern contains many lines but they are not sharp or well defined. The pattern for

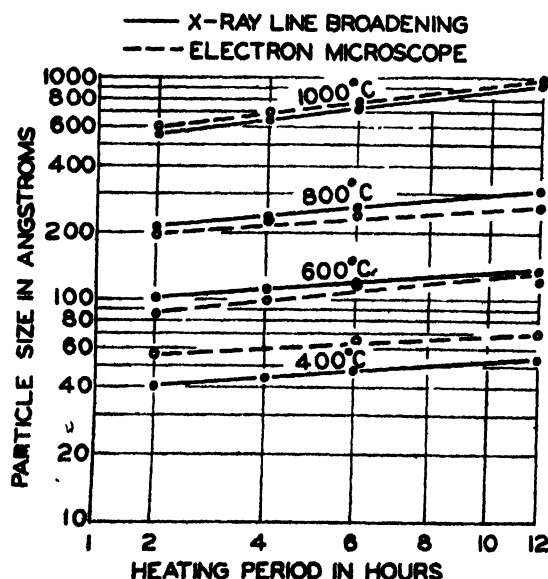


Fig. 7. Particle size in angstroms versus heating period in hours for conversion temperatures from 400°C to 1000°C. Solid lines represent results from x-ray line broadening and dotted lines, results from direct measurement on electron micrographs.

⁹ H. Friedman, Electronics 18, 132 (April 1945).

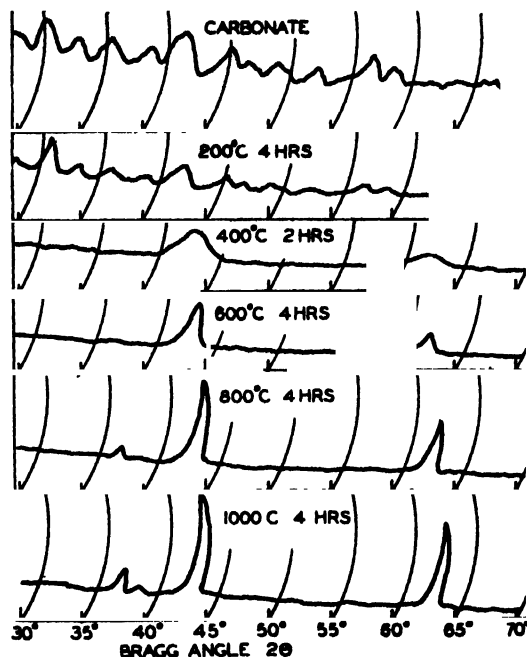


FIG. 8. Automatically recorded x-ray diffraction patterns for magnesium carbonate and magnesium oxide converted at temperatures from 400°C and 1000°C.

200°C is the same as for the untreated carbonate. Close examination reveals that there are no MgO lines present. This confirmed what was observed on the electron micrographs, which showed no change after treatment at 200°C. At 400°C there was no semblance of the carbonate pattern, but two broad halos appeared corresponding in position to the two strongest lines in the MgO pattern. As the temperature was increased, the lines became sharper and more intense until at 1000°C they were almost as sharp as the lines from a powdered quartz crystal. The MgO line at 2θ equals 43° was chosen and the intensity measured at discreet points in this neighborhood by counting for a given time interval. The counts were corrected for non-linearity of Geiger counter response and the lines were plotted as shown in Fig. 9. The width at half-maximum could be measured accurately to $\pm \frac{1}{2}$ minute of arc for the higher temperatures and to ± 2 minutes of arc for the lowest temperatures. The line from quartz at 2θ equals 42.2° was measured to obtain the instrument width used in correcting the broadening.

From the measurements of the width at half-

maximum, the diameter of the particles was calculated using Eq. (1). The results are shown as the solid lines in Fig. 7. The agreement between the two methods was ± 10 percent for all particles larger than 100A.

EFFECT OF NON-UNIFORM PARTICLE SIZE

The above data were for specimens whose distribution curves showed a small range of particle size. The following are some data on mechanical mixtures of specimens of two different sizes. The distribution curve for such mixtures has two peaks, one for each of the components. Figure 10 shows the x-ray results for three specimens, A, 1000°C, 6 hr.; B, 400°C, 12 hr.; and C, a 50-50 mixture of A and B. The line for the 50-50 mixture had the same width at half-maximum as the 1000°C sample indicating the same particle size. However, the peak in-

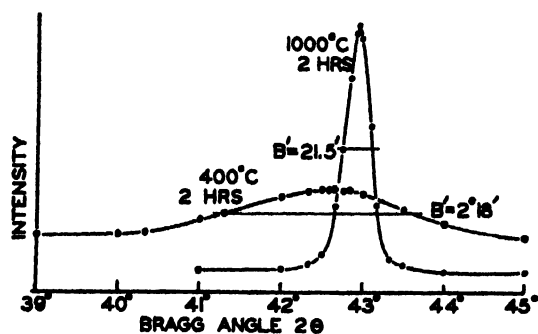


FIG. 9. Typical diffraction lines from magnesium oxide converted from the carbonate at 400°C and 1000°C. The intensity was measured with a Geiger counter at the points shown on the curves.

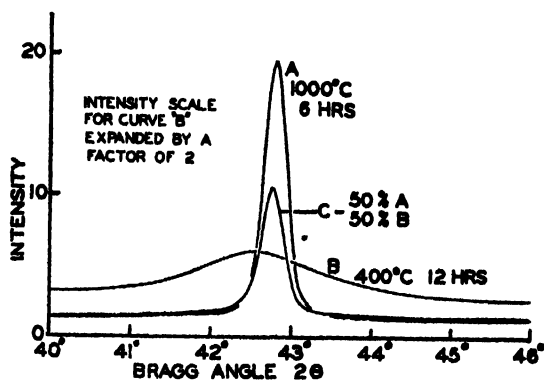


FIG. 10. Effect of a mixture of particles of different sizes. A is for particles approximately 700A in diameter and B is for particles approximately 50A in diameter. C is for a 50-50 mixture by weight of A and B.

tensity was approximately half as great. Thus the maximum intensity could be used as a qualitative measure of the amount of large size material present.

When the size of the two components was not greatly different so that the peaks of the distribution curves for the two were not resolved, the true particle size could not be determined. Figure 11 shows the x-ray lines for three specimens, *D*, 800°C, 6 hr.; *E*, 600°C, 6 hr.; and *F*, a 50-50 mixture of *D* and *E*. The line from the 50-50 mixture is intermediate in both width and height between the two and a measure of the width gives particle size approximately half-way in between the two.

SUMMARY

Electron micrographs, at known magnifications, were used as comparison standards for x-ray line broadening determinations. The line broadening was measured with a focusing type of Geiger counter spectrometer. Results by the two methods agreed to ± 10 percent except for

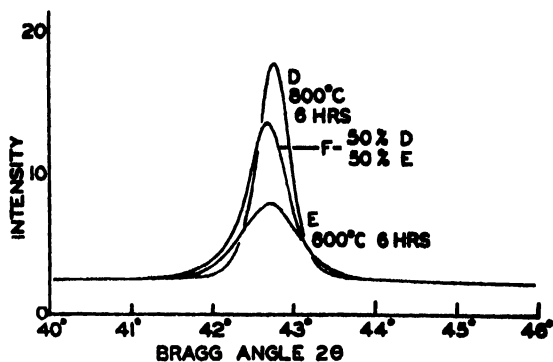


FIG. 11.

particles smaller than 100A where the accuracy with which they could be measured on electron micrographs was only of the order of 25 percent.

Although all the x-ray data shown here were for the magnesium oxide line at $43^\circ 2\theta$, data were also taken for the second strongest line at $62^\circ 2\theta$ and the results agreed to within 10 percent. The material for this paper was condensed from N.R.L. report H-2434 prepared in the summer of 1944.

The Effect of Hydrostatic Pressure on Plastic Flow under Shearing Stress

P. W. BRIDGMAN

The Physics Laboratories, Harvard University, Boston, Massachusetts

(Received April 25, 1946)

This paper presents measurements of the forces required to drive a punch into plates of several grades of steel as a function of the distance of penetration, the entire punching operation being conducted in a fluid medium subjected to hydrostatic pressures up to 30,000 kg/cm². Qualitatively, the effects are similar to those already found for the tensile properties of steel subjected to hydrostatic pressure, namely, ductility is greatly increased, and greatly increased distortion is tolerated without fracture. The effects are similar in general character for all the steels experimented on here, but are quantitatively accentuated for the softer steels. At a pressure of 20,000 kg/cm² or more a punch may be driven completely through a plate of mild steel, with no loss of coherence at any stage of the process, and with strain hardening, when expressed in terms of true shearing stress, which may increase by a factor of as

much as 3. If the punching operation is suspended at any intermediate stage before complete penetration and afterward completed at atmospheric pressure, very material strengthening will be found as compared to virgin material of the same geometrical configuration. There are certain qualitative differences in the details of the ductility exhibited during the processes of punching and pulling. At intermediate pressures and penetrations the true shearing stress in punching may, under proper circumstances, exhibit maxima, for which there is no analog in the pulling operation. This is to be understood in terms of a difference of geometry, there being greater opportunity for self-healing during the punching operation, and, furthermore, partial deterioration of the coherence of the metal not necessarily leading to complete catastrophe.

INTRODUCTION

A NUMBER of experiments have established the very large effect of pressures up to 30,000 kg/cm² in increasing ductility under ten-

sile stresses.¹ In this paper a study is made of the

¹ P. W. Bridgman, *Rev. Mod. Phys.* 17, 3-14, 1945. Eight reports to the Watertown Arsenal, March 1943, through Dec. 1944, numbered No. WAL 111/7, 111/7-1, etc., through 111/7-7.

effect of pressure in increasing the magnitude of the flow which can be tolerated without fracture under the action of shearing stresses. In general, the results are of the same qualitative character as for tensile stress, in that much greater deformations are possible without fracture under pressure than at atmospheric pressure.

THE METHOD

Perhaps the most natural method would be that of twisting a rod. There are, however, great technical difficulties in applying, inside a vessel carrying high hydrostatic pressure, rotations of the magnitude which experiments with simple compressive stresses¹ suggested would be necessary to produce fracture, and no attempt was made to develop this method. Some form of punching seemed indicated by the instrumental limitations. Preliminary experiments were made in which a rod was passed through holes in a cylindrical sleeve and a closely fitting cylindrical core and sheared through by longitudinal displacement of the core within the sleeve. The apparatus was adapted to materials of various strengths by providing holes by which several rods could be sheared simultaneously. There were, however, difficulties in so fitting the specimens to the holes as to start the plastic flow in all specimens simultaneously, and the method was abandoned.

The method finally adopted was the punching of a circular disk; the apparatus is shown in Fig. 1. The specimen to be punched, shown shaded, is compressed tightly between an upper and a lower annular clamp in order to prevent warping when the punch penetrates. The disk to

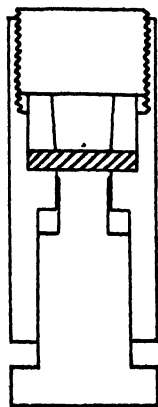


FIG. 1. The punch with specimen, shaded, in position.

¹ P. W. Bridgman, *J. App. Phys.* 14, 273-283 (1943).

be punched is 0.050 or 0.03 inch thick, depending on its hardness. The two surfaces were ground to parallelism in a surface grinder. The punch was of "Teton" tool steel, left glass hard. The edge of the punch was sharp, and the face flat. The cylindrical side was full diameter for a length of 0.019 inch and then relieved. The die through which the punching is expelled was of the same Teton steel, with sharp edges, and tapered at a double angle of 1° to eliminate friction on the expelled punching in the die. The cylindrical holder for punch, specimen, and die was of a Cr Mo steel, heat treated to Rockwell C hardness 30. All parts were ground and fitted accurately together. The specimen in the holder was mounted in the 30,000 apparatus in the same position as that taken previously by the tension specimens. The force driving the punch was measured with the same "grid" that was used to measure tensile load, the displacement of the punch was measured in the same way as the elongation of the tension specimens, and the whole technique of the experiment was very similar. Since the total motion of the punch from initial contact to complete penetration was less than the total elongation of the tension specimens, the total change of hydrostatic pressure during the course of the punching was less than in the corresponding tension experiment, which is an advantage.

The measurements consist in simultaneous measurements of the hydrostatic pressure, the force on the punch, and the displacement of the punch, the latter being measured externally from the displacement of the piston producing the hydrostatic pressure. This latter displacement includes, in addition to the actual penetration of the punch, the elastic distortions of all the intermediate parts, including the punch itself, the transmitting parts within the pressure chamber, and the external piston. These elastic distortions were corrected out of the final result by making a sufficient number of readings before plastic flow began to establish the slope of the line of elastic distortion, and subtracting from the total displacement after plastic flow began an amount proportional to the total load and the slope.

The total load on the punch was converted into shearing stress per unit original area by dividing the total load, after correction for

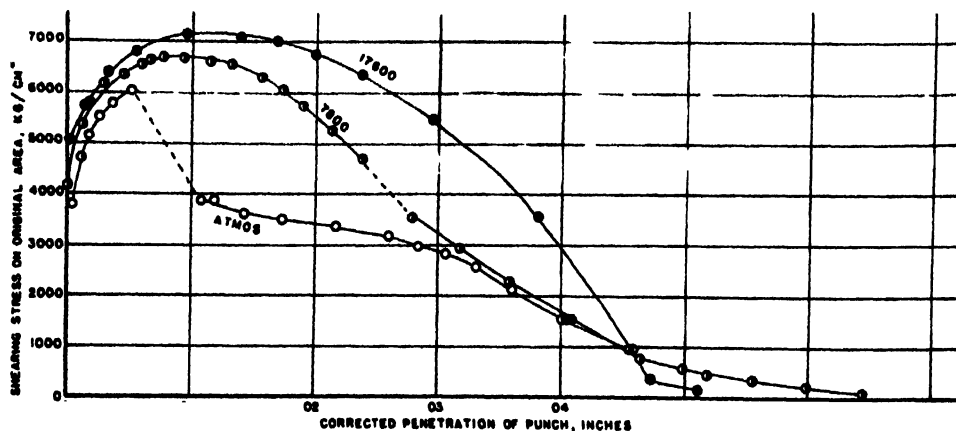


FIG. 2. The experimental data at several different pressures, indicated on the curves, for heat treatment No. 3 of steel No. 4.

friction on the punch, by the area of the sheared region (product of thickness of plate and circumference of punch). This shearing stress on the original area was then converted to "true" shearing stress by multiplying by the ratio of the initial thickness to the running thickness. The "true" shearing stress was finally plotted against the penetration of the punch. The friction on the punch was computed from the readings at the end of the punching process, when the face of the punch had been driven completely through the plate, but the unrelieved sides of the punch were still in contact with the plate. This contact persisted for a displacement of 0.019 inch; during contact the driving force dropped proportionally to the distance still to go. The maximum correction for the friction of the punch was about 10 percent.

It would have been desirable if the "true" shearing stress could have been plotted against the strain, so as to get a strain hardening curve for pure shear. The conditions of plastic flow were not, however, well enough defined to permit any clean cut evaluation of the strain. Ideally the strain might be localized in the thin cylindrical shell included between the diameter of the punch and the diameter of the die. Actually, however, the strain spread out into the surrounding region from this cylindrical shell in a way that would be most difficult to compute. The question was investigated by using punches of different diameters, thus varying the thickness of the cylindrical shell. If a true strain-hardening curve of stress against shearing strain could be obtained by the idealized calculation, then there should be a simple connection between the

results obtained with the different punches, the true shearing stress for a certain penetration with one punch being the same as that at a penetration twice as great with a punch of twice the clearance. Three punches were used, with clearances varying from 0.0040 to 0.0008 inch. At the same penetration of the punch the force with the largest clearance was only 3 or 4 percent less than with the smallest clearance. This shows that the plastic flow must be pretty well spread outside the idealized cylindrical shell, and frustrates the attempt to express the results in terms of absolute strains. For this reason these experiments on punching have a more qualitative significance than the tension experiments.

THE MEASUREMENTS

All the final measurements were made with the largest punch, with a clearance of 0.0008 inch. Measurements were made on five grades of steel. The first was a 1045 steel from the Watertown Arsenal in the "as received" condition; tensile data for this will be found in my reports to the Arsenal.³ The others were two steels, Nos. 2 and 4, which had been used in certain experiments for the N.D.R.C. during the war,⁴ in two different heat treatments. The compositions and heat treatments were as follows:

Steel No. 2, C 0.45, Mn 0.83, P 0.016, S 0.035, Si 0.19.
Steel No. 4, C 0.90, Mn 0.47, P 0.015, S 0.036, Si 0.11.
Heat treatment No. 3; annealed at 1900°F for 0.5 hour, giving a coarse grained structure.
Heat treatment No. 9; held at 1800°F for 0.25 hour, quenched into brine, and drawn at 600°F for 1 hour.

³ In particular, in Report No. WAL 111/7.

⁴ These reports have been recently declassified (7/12/46).

The qualitative nature of the effects is shown in Fig. 2, in which is plotted for heat treatment No. 4 the shearing stress on the original area (which is proportional to the total driving force on the punch) against the corrected penetration of the punch. In this instance the material "breaks" at atmospheric pressure after a penetration of the punch to one-tenth the thickness of the plate. This is indicated by the dotted line in Fig. 2. The force jumps down, and the punch jumps in, the intermediate configurations being unstable and controlled by elastic distortion in the punch itself and the transmitting members. After the break, a force comparable with the maximum is still required to advance the punch, the force gradually dropping off as penetration becomes complete. The first effect of hydrostatic pressure is to suppress the break, displacing it at the same time toward higher penetrations. In Fig. 2, at the next pressure above atmospheric, 7800 kg/cm², only a vestige of the break remains, and it has been displaced from a penetration of 0.005 to 0.025 inch (one-half the thickness of the

plate). At the next higher pressure, 17,000, no vestige of the break remains.

On converting total loads into true shearing stresses and correcting for friction, Fig. 2 is transformed into Fig. 3. The secondary rise in the true stress at atmospheric pressure after passing the break is explained by the character of the punching. This is not truly cylindrical, but the exit diameter, in contact with the die, is slightly less than the diameter at the face of the punch, so that the punching has the shape of a truncated cone. After the break, the larger base of the cone has to be forced through a smaller aperture, like forcing a conical rubber stopper the "wrong" way through an aperture, thus accounting for the increase in force. The drop of force when expulsion nears completion is caused by distortion in the thin lip of the plate. With increase of hydrostatic pressure the truncated cone becomes a true cylinder, brightly burnished over its entire external surface, instead of being burnished only over the ring corresponding to the penetration of the punch before the break. The transition between the two conditions is gradual.

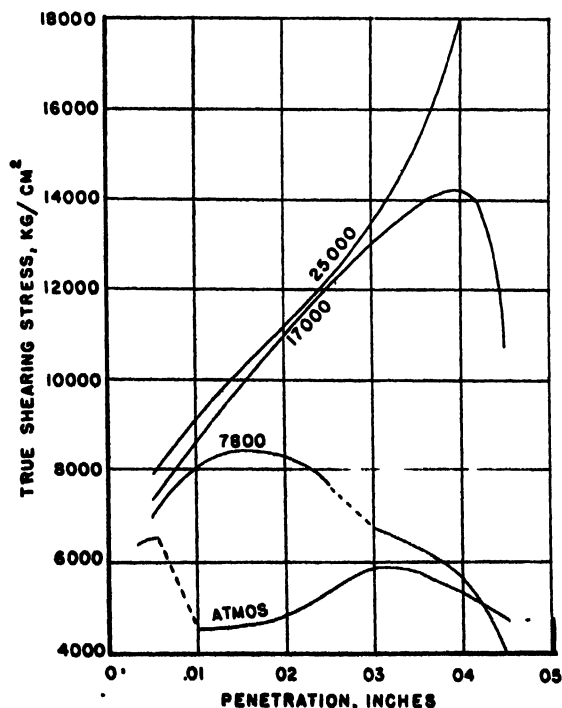


FIG. 3. The "true" shearing stresses corresponding to the data shown in Fig. 2. The original data for the run at 25,000 kg/cm² are not shown in Fig. 2, in order to avoid confusion with the data for 17,600.

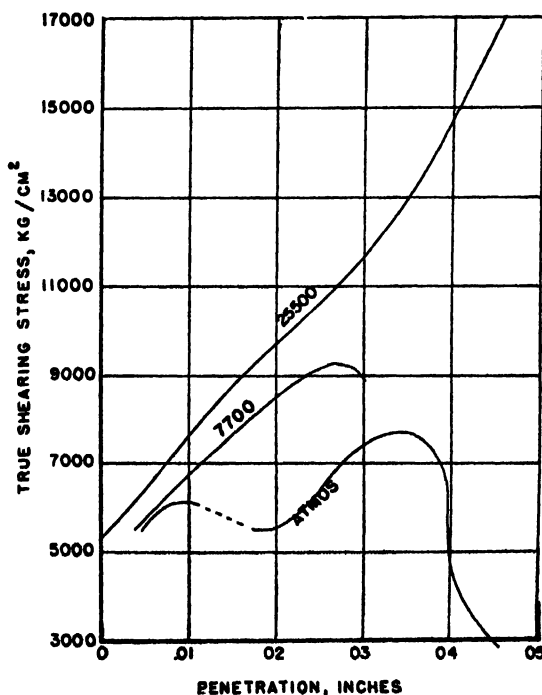


FIG. 4. "True" shearing stress versus penetration at several pressures for the 1045 steel, "as received." The heavy line at .05 inch indicates the total thickness of the plate.

In Fig. 3 the true stress at the first pressure above atmospheric, 7800, rises to a maximum and then drops slightly before the vestigial break occurs. That is, there has been some deterioration of the steel after a certain degree of penetration, but no wide open break. At the next pressure, 17,400, strain hardening is approximately linear with the penetration up to a penetration to 0.8 the original thickness. The drop beyond 0.8 represents deterioration, without actual fracture. The upper curve in Fig. 3 is for a pressure of 25,000; the corresponding curve was not drawn in Fig. 2. At this pressure any deterioration is displaced so nearly to complete penetration as to be beyond experimental reach. The upward curvature at the end of this curve is probably real, and corresponds to the fact that the average effective strain increases at an accelerated rate as penetration nears completion.

The Watertown 1045 steel and heat treatment No. 3 for steels No. 2 are much similar. In Fig. 4 is shown true stress *versus* penetration for the 1045 steel, and in Fig. 5 for heat treatment No. 3

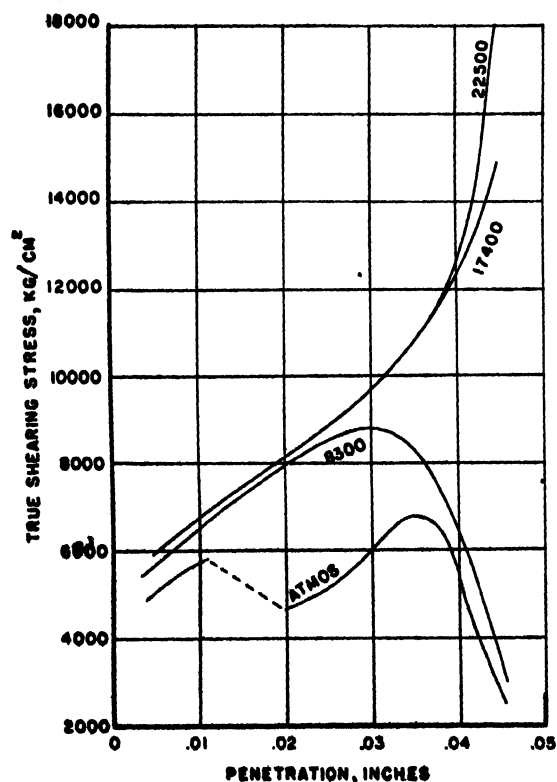


FIG. 5. "True" shearing stress *versus* penetration at several pressures for heat treatment No. 3 of steel No. 2.

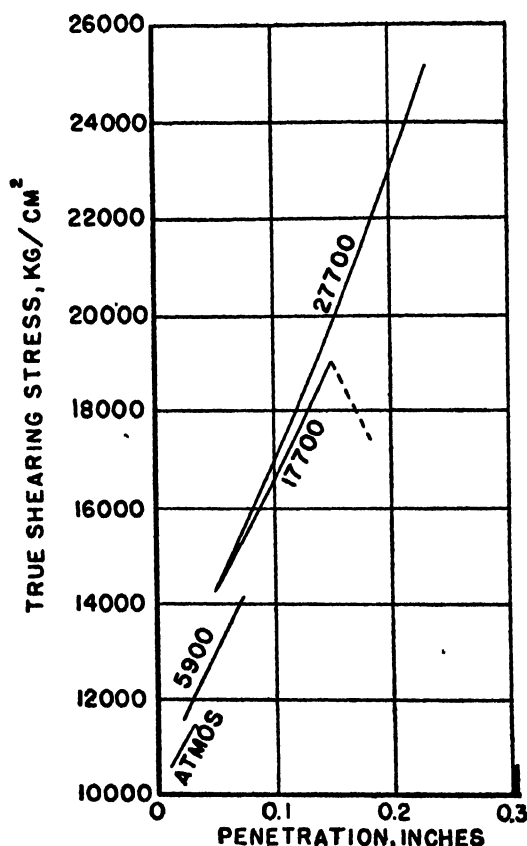


FIG. 6. "True" shearing stress *versus* penetration at several pressures for heat treatment No. 9 of steel No. 2.

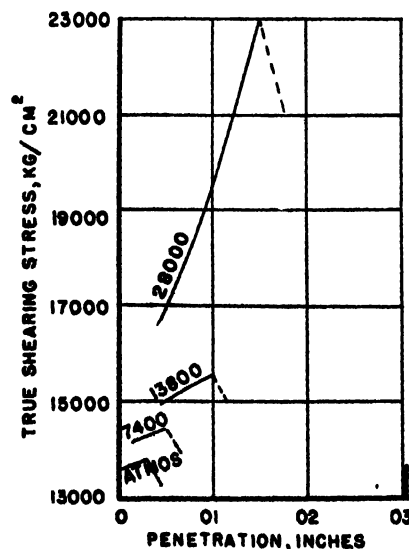


FIG. 7. "True" shearing stress *versus* penetration at several pressures for heat treatment No. 9 of steel No. 4. The total thickness of the plate was .0302 inch.

of steel No. 2. These steels are both softer than No. 4, and the curves for both show upward curvature at the deepest penetrations and the highest pressure. Toward the end of penetration, the strain is increasing more rapidly than in the initial stages, because it has less chance to spread out from the idealized cylindrical shell, and the steels are soft enough not to show the deterioration at high strains shown by No. 4.

Figures 6 and 7 show the results for heat treatment No. 9 on steels No. 2 and No. 4. This heat treatment gives a much harder product, with higher resistance to punching and much less plastic yield when the break does occur. The effects of pressure here are in the same direction qualitatively as for the softer steels, but they do not go nearly so far. The fracture phenomena are not suppressed in the pressure range shown in the figures; however, the penetration of the punch before the break occurs is much increased by pressure. This agrees with the results previously found for tension, namely that the phenomena for the harder steels are qualitatively the same as for the softer steels, but the scale of the effects is changed, a much higher pressure being required for the same plastic flow without fracture in a harder steel as compared with a softer steel.

In order to demonstrate that the punching may be moved bodily through the surrounding metal under pressure without fracture, an experiment was made on the 1045 steel in which the punch was driven 0.9 through the plate at a pressure of 23,000 kg/cm², and then the punching operation was completed at atmospheric pressure. For comparison, another plate was punched at atmospheric pressure which had been machined so as to have in it a depression of the same dimensions as that which had been driven into the plate by the punch under 23,000. At atmospheric pressure the maximum force required to start the punch moving through the prepunched plate was more than 80 percent greater than the maximum force supported by the geometrically similar virgin specimen. This is evidence of considerable strain hardening, and incidentally shows that the prepunched plate could not have had many wide open fractures, if any.

In another similar experiment the punch was pushed 0.4 through the plate at a pressure of

26,000 kg/cm², and the punching operation was then completed at atmospheric pressure. In this case 42 percent more force was required to start the punch moving through the prepunched plate than was needed for a geometrically similar virgin specimen. The prepunched plate behaved in the way usual for strain hardened material in that the maximum penetration of the punch up to the break (calculated from the new zero) was only 0.38 of that for the virgin specimen. The true shearing stress at the maximum penetration under pressure was 8900 kg/cm²; it required 7300 true stress to make flow resume at atmospheric pressure. Some of the difference between these two figures may be caused by a loss of strain hardening by an annealing effect on resting. However, if the total difference between these two figures is ascribed to a specific effect of hydrostatic pressure, it would mean a raising of the plastic flow stress by only 8 percent per 10,000 kg/cm² pressure, and this is slightly less than has been previously found under similar circumstances.

DISCUSSION

The qualitative resemblance of the effects of pressure on plastic flow in tension and in shear has already been commented on. The greatest increase of shearing stress observed is that for the 1045 steel, which shows a maximum increase under 26,000 kg/cm² of something more than threefold. In tension a threefold raising of the flow stress for this steel is produced by a strain of 3.75, or an elongation of 43-fold (second Arsenal report).

In addition to this rough qualitative resemblance there is one significant qualitative difference between the strain-hardening curves for tension and for pure shear. The strain-hardening curve for tension rises linearly with increasing strain until the point of fracture is reached, which occurs abruptly, with no previous warning. The difference between a strain-hardening curve for tension at one pressure and another higher pressure is merely, to a first approximation, that rupture on the second curve is not reached until a higher strain than on the first. In shear, however, the strain-hardening curves for different pressures do not coincide over their entire extent, but, especially at the lower pressures, the

approaching fracture announces itself by a dropping away from the strain-hardening curve for high pressures. In a certain range of pressure the strain-hardening curve may show a maximum. This is evident in Figs. 3-5 on the curves for the first pressure above atmospheric. This would indicate a deterioration of the steel before complete fracture. The physical explanation of the difference between shearing and tension is to be sought in the difference of the stability relations. If an incipient fracture occurs in tension the adjacent parts are immediately removed

from the region of atomic interaction and the fracture is catastrophic. On the other hand, when tangential slip of one part on another occurs, the slipping parts are not necessarily removed from atomic contact and cohesion remains to an extent depending on the degree of destruction of the lattice structure. In other words, in tangential slip the phenomenon of self-healing can occur, but it cannot in tension. Hydrostatic pressure keeps the slipping parts in contact, so that the self-healing is more complete the higher the pressure.

Hysteresis and Related Elastic Properties of Tire Cords

HELMUT WAKEHAM AND EDITH HONOLD

Southern Regional Research Laboratory, New Orleans, Louisiana*

(Received May 25, 1946)

A method of evaluating mechanical hysteresis, elastic modulus, elongation, and growth rate of a yarn or cord subjected to cyclic loads in a fatigue test is described. The test is applied to seven typical tire cords—one Nylon, two rayon, and four cotton cords including both high stretch and low stretch cords. Results are presented showing the effect of cycle number, moisture content, temperature, and load range on the elastic properties of each of the cords tested. The influence of cyclic loading on the stress-strain curve is also discussed. It is concluded that moisture content is the dominant factor influencing the elastic properties of tire cords and that the heat produced by mechanical hysteresis of the tire cord contributes appreciably to the heat build-up observed in tires during heavy duty service.

IN a previous publication¹ preliminary data on the mechanical hysteresis of cotton and rayon tire cords were presented. The results of a more extended investigation of the hysteresis and related elastic properties of tire cords are reported in the present paper.

The phenomenon of elastic hysteresis for textile materials has received little attention by previous investigators in the field of textile properties. Barker and Tunstall² obtained extension and recovery curves for wool fibers by means of an autographic recording stress-strain apparatus in which the rate of loading and unloading was relatively slow. Later methods of investigation were of the same type and ex-

tended over only a few cycles.³⁻⁵ Such published results are of limited applicability, however, because they give little insight into the properties of the material during prolonged use. Whereas exposed fabric, such as wearing apparel, usually can be reliably tested at convenient intervals during its lifetime, the evaluation of the properties of tire cords during service presents a unique problem because the cord is vulcanized in rubber and cannot be removed without destroying the usefulness of the tire. There is need for a testing procedure in which the change in properties during mechanical deterioration of the cord to the point of failure will parallel the changes in actual service.

In a general sense, any test in which the test

* One of the laboratories of the Bureau of Agricultural and Industrial Chemistry, Agricultural Research Administration, U. S. Department of Agriculture.

¹ H. Wakeham, E. Honold, and E. L. Skau, *J. App. Phys.* 16, 388 (1945).

² S. G. Barker and N. Tunstall, *Trans. Faraday Soc.* 25, 103 (1929).

³ W. F. Busse, E. T. Lessig, D. L. Loughborough, and L. Larrick, *J. App. Phys.* 13, 715 (1942).

⁴ W. S. Denham and T. Lonsdale, *Trans. Faraday Soc.* 29, 305 (1933).

⁵ O. Eisenhut and W. Grether, *Melliand Textilber.* 22, 122 (1941).

specimen is subjected to cyclic deformation is a fatigue test. Numerous tests of this type for tire cord both in and out of the tire have been described^{6,7} in which the time, or number of cycles, to fatigue failure is measured. No doubt the information so derived is extremely useful, especially in control testing, but the research worker interested in improving cord needs to find out also what changes take place in the cord during its deterioration to failure. This phase of the problem seems to have been almost completely overlooked by workers in the tire cord field.

One of the major problems in the design of large heavy-duty truck tires has been the high temperatures built up during service as a result of the imperfect elasticity of the tire. The heat generated is unquestionably related to the elastic hysteresis of the rubber and the tire cord. While rubber hysteresis has received considerable attention by investigators in the rubber industries,⁸ relatively little is known of cord hysteresis and the contribution which it makes to the high temperatures observed in tires during service.

It was with these facts in mind that the comparison of elastic properties described below was undertaken.

APPARATUS

When the tension on a cord is progressively increased to a maximum and then lowered again, the elongation for a given load differs, depending upon whether the tension is increasing or decreasing. If percent elongation of the cord is plotted against the increasing and decreasing loads, a loop as shown in Fig. 2 (right-hand side) is obtained. The area within the loop is directly related to energy loss or hysteresis, because of imperfect elasticity of the cord. In the apparatus used the cord is stretched and relaxed in a straight line while the record of the elongation and load are traced on the moving paper of a kymograph. The tracings so obtained are sinusoidal-like curves of different amplitudes from which the hysteresis and other elastic properties may be calculated.

⁶ W. H. Bradshaw, A.S.T.M. Bull. 13 (October, 1945).

⁷ L. Larrick, Textile World, 95[5], 107 (1945).

⁸ See for example: J. H. Dillon, I. B. Prettyman, and G. L. Hall, J. App. Phys. 15, 309 (1944). E. T. Lessig, Ind. Eng. Chem. Anal. Ed. 9, 582 (1937). M. Mooney and R. H. Gerke, Ind. Rubber World, 103[4], 29 (1941). H. S. Sack, J. Motz, and R. N. Work, J. App. Phys. 15, 396 (1944).

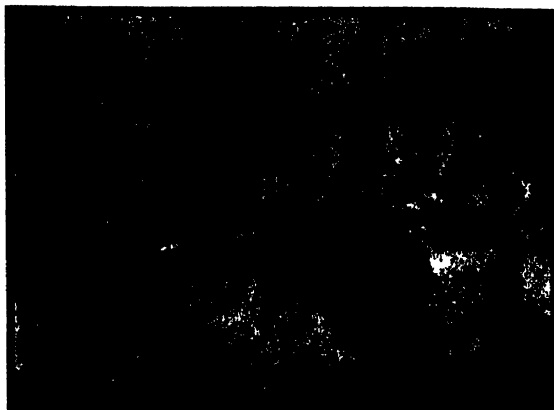


FIG. 1. Mechanical loading and recording apparatus at the left-hand end of the oven. See text for complete description.

In this group of experiments, the cord is tested within a specially constructed oven, as described in the previous paper,¹ but the arrangement of the apparatus beyond each end of the oven (o, Fig. 1) has been changed. In series, starting at the left-hand end, there are motor-driven geared eccentric (a), a calibrated spring (b), the left-hand tie, the cord, the right-hand tie, a connection, and movable brace which permits growth take-up. The eccentric and spring apply the increasing and decreasing load cycle to the cord; a slotted insert (c), which moves diametrically across the face of the eccentric and locks into position, permits the adjustment of the stroke to the selected load range. The connection between this pegged insert and the spring is made by means of a rigid bar (d) and a heavy wire (e). The latter is always in line with the cord, being held in position by frictionless pulleys. At each end of the spring, a recording pen is attached (f, g). Rigidly and immovably fastened to the framework is a third pen (h) which draws a straight line on the kymograph roll to serve as a guide line from which all calculations are made.

The pens rest lightly on paper passing through the kymograph (i), and trace the above-mentioned simultaneous curves so that at any instant during a cycle, the distance between the pens is convertible to load and the tracing of the right-hand pen (g) permits the calculation of percent elongation of the cord. As a cord is taken through successive cycles, a permanent non-recoverable stretch occurs, which must be taken

up to keep the pens on the paper while maintaining the selected load range. This is accomplished by an arrangement beyond the right-hand end of the oven (o). The right tie extends from the cord through a connector to a ring which slips around the shaft of a movable brace, the position of which is changed as the cord stretches. From suitable measurements on the tie at each end the length of the cord can be calculated.

The apparatus at each end of the oven is mounted on a rigid frame so that the observed motion of the elongation pen accurately represents the stretch in the cord rather than any flexibility or relative motion in the system. The precision of this arrangement was tested by placing a tempered steel wire of B. & S. No. 30 gauge in the position to be occupied by the cord and its ties. The load cycle was applied to the wire and the elongation recorded was found to be closely comparable with that calculated for a steel wire by known values of Young's modulus for steel.

PROCEDURE

The cord, of approximately 100-cm length, is taken through the cycles at a rate of 1 cycle per second. Tracings are made at designated cycle intervals as these are indicated by an automatic counter attached to the apparatus. Temperatures in the tubular oven are controlled and measured in the manner previously described.¹

The moisture contents of the cords, in percent of their dry weights (frequently referred to as moisture "regains") are found by actual determination of the moisture after each run is completed. The standard procedure for obtaining moisture and oven dry weight is followed throughout.² In certain experiments the cords to which the cyclic loads had been applied were later tested for their stress-strain characteristics. Since in such cases it was impossible to determine the moisture directly, the moisture content of another length of the same cord simultaneously contained in the tube was taken as the content of the loaded cord. Check determinations in which the moisture contents of both loaded and dummy cords were determined agreed within the experimental error of an individual measurement.

¹A.S.T.M. Standards, Part III, D269-42T, p. 1952 (1944).

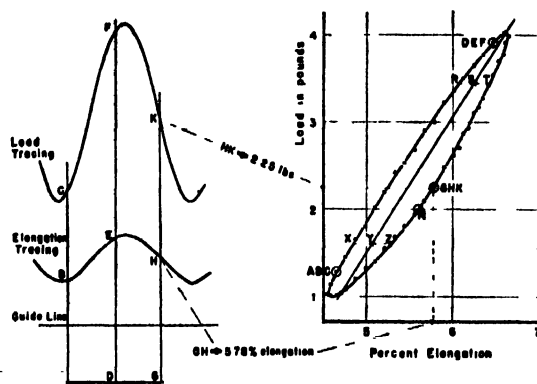


FIG. 2. Method of calculating hysteresis loop on the right from the kymograph tracings on the left. For complete details of this calculation see text.

The moisture content of the cord in the test oven is controlled by preconditioning the cord and by passing conditioned air through the tube while the run is in progress. The apparatus is in a room in which the humidity is controlled to ± 3 percent and cord moisture over the middle of the range is changed by merely adjusting the room humidity. For the dry runs the cord is first dried in a vacuum before insertion into the tubular oven and then kept surrounded by dry air during the run. When it is desired to keep the cord with a definite moisture content at high temperatures the air passed through the tube is conditioned by first being bubbled through water in a constant temperature bath.

At the start of a run, the cord is introduced into the oven, a 4-ounce weight attached, and measurements taken to establish the cord length which is used as a basis of all percent elongation calculations. After all necessary connections and preliminary adjustments are made, the run is started but may be interrupted briefly at any time to make further adjustments as the cord grows in length.

Figure 2 depicts schematically the method of plotting the hysteresis loop from the kymograph tracings. The base line is set parallel to and at a distance above or below the guide line as dictated by the observed length of the cord at a particular cycle. Two transparent right triangles (similar to those used in drafting work) are used to facilitate the measurements. The perpendicular side of the one is ruled in millimeters and that of the other in pound-equivalent distances. By fitting the base of the millimeter triangle against

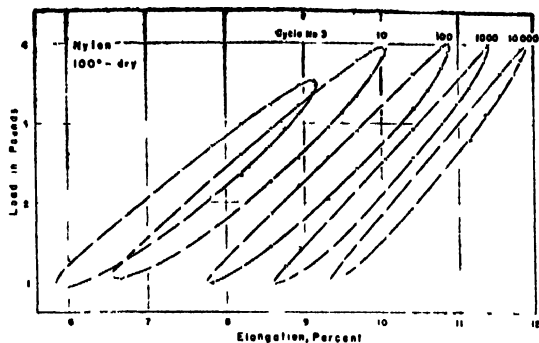


FIG. 3. Typical mechanical hysteresis loops at various cycle numbers for dry Nylon at 100°C.

the base line the distance to the elongation tracing (AB) can be read directly in millimeters and converted into percent elongation. The ruled side of the load triangle is moved along against the millimeter scale to convert the distance between the tracings (BC) to load in pounds. These simultaneous readings are plotted on coordinate paper as shown—thus values from AB and BC are used to plot point ABC , values from DE and EF to plot point DEF , etc. Sufficient pairs of readings are made to permit drawing of the hysteresis loop.

From the hysteresis loop elastic characteristics of the cord are obtained as follows.

Hysteresis

The loops are integrated graphically with a planimeter and converted into inch-pound values per 100 inches of length. Hysteresis may be expressed in any energy units as, for example, calories per centimeter length of cord or calories per gram of cord.

$$1 \text{ inch-pound}/100 \text{ inches} = 1.06 \times 10^{-4} \text{ calorie/centimeter.}$$

$$\text{Calories/gram} = \frac{(\text{calories/centimeter}) \times 10^6}{\text{grec}}$$

In the results reported below the hysteresis is expressed in inch-pound per 100 inches because this value is obtained directly from the plots of load in pounds *vs.* percent elongation.

Elastic Modulus

At each of two load levels, arbitrarily chosen, of 0.6 pound from top and bottom of each loop (RT and XZ , respectively), the average of the

percent elongation readings is marked (S and I') and a line drawn through these two points to establish the slope of the loop. Elastic modulus is determined as dynes per square centimeter from the formula derived from Young's modulus:

$$M = t \times \frac{100}{3.1416g^2/4} \times 6.895 \times 10^4,$$

in which t is the tangent value of line SI' , g is the gauge or diameter of the cord in inches, and 6.895×10^4 is the factor converting pounds per square inch into dynes per square centimeter.

Percent Elongation

The percent elongation at 2 pounds during the unloading portion of the cycle is arbitrarily chosen to represent the elongation of the cord for that cycle. This corresponds to point N in Fig. 2.

Rate of Growth

This is defined as the increase in cord elongation produced by a tenfold increase in cycle number as, for example, the difference between the elongations for the tenth and for the hundredth cycle.

Hysteresis loops were plotted at selected intervals during the run; i.e., at cycle numbers 3, 6, 10, 30, 100, 300, etc., for 10,000 cycles. Such a series of loops for dry Nylon at 100°C is shown in Fig. 3. Secondary graphs were made of hysteresis, elastic modulus, and elongation, each *vs.* cycle number on a logarithmic scale for convenience in graphing. These are shown in the results described below.

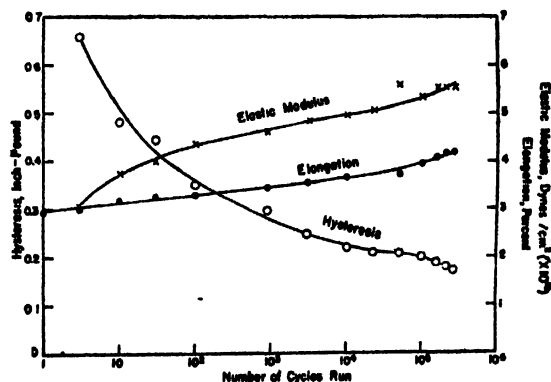


FIG. 4. Hysteresis, elastic modulus, and elongation *vs.* cycle number to point of rupture for dry experimental cotton cord-2 at 145°C.

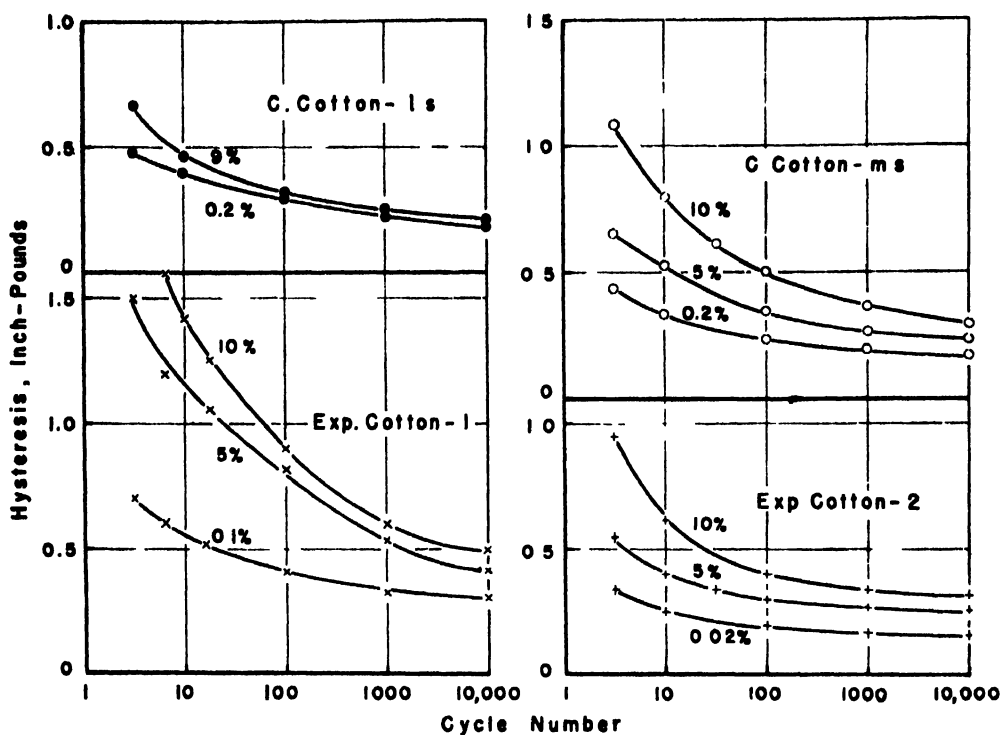


FIG. 5. Hysteresis vs. cycle number at constant moisture and 25°C—cotton cords.

EXPERIMENTAL RESULTS

Seven different cords were chosen for this study: (1) 210/4/2 Nylon; (2) 1100/2 viscose rayon; (3) 2200/2 viscous rayon; (4) 18/4/3 medium stretch cotton; (5) 16/4/3 low stretch

cotton; (6) 17/4/3 unstretched experimental cotton; and (7) 17/4/3 stretched experimental cotton. Some physical properties of these cords under A.S.T.M. standard testing conditions¹⁰ are shown in Table I. The two rayon cords have

TABLE I. Some physical properties of tire cords used in study of elastic properties

Cord designation	Nylon		Rayon		Cotton		Experimental	
			1	2	Commercial Medium stretch	Low stretch	1	2
Gauge	0.016		0.023	0.033	0.031	0.032	0.035	0.031
Construction ^a	210/4/2		1100/2	2200/3	18/4/3	16/4/3	17/4/3	17/4/3
Grex (dry) ^b	1986		2444	5366	4342	5064	4716	4446
Standard conditions	Breaking load	In lb.	25.2	14.2	28.3	20.0	26.7	23.4
		In g/grex	5.76	2.64	2.39	2.09	2.39	2.25
	Percent elongation	At 10 lb.	12.4	9.8	6.1	7.0	5.0	14.7
		At break	22.0	16.6	17.5	12.3	11.2	21.8
• ^c	Moisture, % of dry weight		4	14	14	8	8	8
Bone dry	Breaking load	In lb.	25.4	18.0	34.3	17.7	24.8	18.5
		In g/grex	5.80	3.34	2.90	1.85	2.22	1.78
	Percent elongation	At 10 lb.	11.8	4.5	3.3	6.0	5.2	10.4
		At break	22.0	10.8	12.2	8.7	9.1	15.4

^a The first figure represents size of single yarns (denier for Nylon and rayon and yarn number for cotton); the second figure represents number of single yarns per ply; and the third figure represents number of plies in the cord.

^b The grex unit is defined as the weight in grams of 10,000 meters of cord. See A. G. Scroggie, Rayon Text. Mo. 25, 45 (1944).

¹⁰ A.S.T.M. Standards, Part III, D179-42, p. 811 (1944).

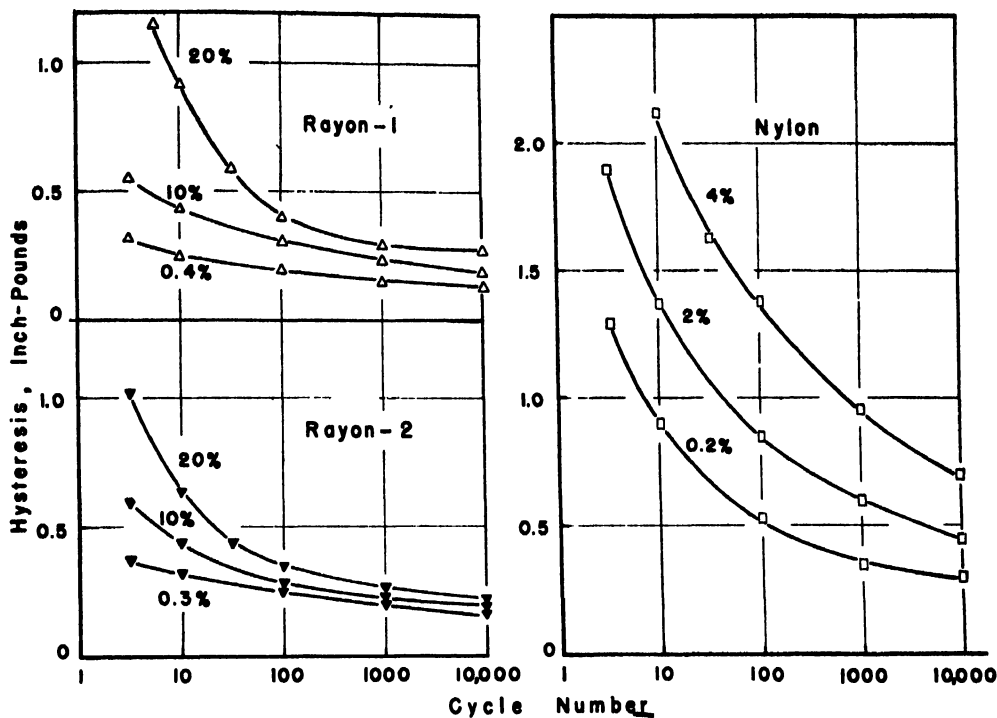


FIG. 6. Hysteresis vs. cycle number at constant moisture and 25°C -rayon and Nylon cords.

weight and strength ratios of approximately 2 :1- and comparisons of these two cords when tested at equal load ranges may be expected to demonstrate to some extent the effect of load. The first two cotton cords are samples of commercial cords which have been used as fabric in regular tires. The last two are experimental cords made in this laboratory as part of the research program on cotton tire cords.

The hysteresis and related properties of a tire cord are functions of many variables such as load range, cycle number, cycle frequency, temperature, and moisture content. In much of this investigation the cyclic tension was kept in the range between 1 to 4 pounds. The cycle frequency was standardized at approximately one cycle per second. Even at this rate 3 hours were required to run off 10,000 cycles. Since it was also desired to investigate the variables of temperature and moisture content, it was necessary to limit the number of cycles run in a routine test to approximately 10,000. Continued operation beyond this point gave little additional information as the test described in Section 1 demonstrated.

It was, furthermore, not practicable to make a

large enough number of runs under identical conditions with samples from any one cord to obtain anything like a statistical average. Duplicate and triplicate runs were made on numerous samples; and in most cases the deviations between individual curves of duplicate runs were less than the differences found between any two cords. Thus, the results herein presented demonstrate real dissimilarities between the cords studied, although the accuracy with which these have been determined leaves considerable room for improvement.

1. Effect of Cycle Number

As the tire cord is subjected to alternating loads its elastic properties "undergo gradual changes. These changes are illustrated in Fig. 4 in which the hysteresis, elastic modulus, and elongation are plotted vs. cycle number (on a logarithmic scale) to the point of failure for experimental cotton cord-2. These curves are typical of those obtained for all cotton, rayon, and Nylon cords tested. In all cases hysteresis decreases whereas elastic modulus and elongation increase with continued cyclic loading.

In the work here reported no study has been

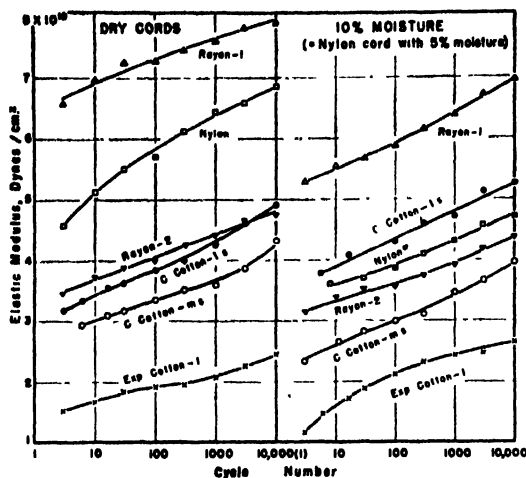


FIG. 7. Elastic modulus vs. cycle number at constant moisture and 25°C.

made of cord properties at or immediately prior to rupture. This has not been practicable because of the long time required to break a cord under the rates and magnitudes of cyclic loads imposed by the apparatus used. There seems little doubt, however, that such a study would throw considerable light on the mechanism of fatigue failure of tire cords and other industrial fabrics.

2. Effect of Moisture

Hysteresis at 25°C as a function of cycle number is shown for each tire cord at various moisture contents in Figs. 5 and 6. It is evident that hysteresis in all cases increases with increase in moisture, with much more increase for Nylon than for the other cords. Hysteresis curves for the low stretch commercial cotton cord are almost independent of moisture content—curves obtained with 4, 5, and 8 percent plotted practically on top of those shown. On the basis of comparable moisture contents the rayon hysteresis is less affected by moisture than that for cotton. Any advantage rayon may possess over cotton in this respect, however, is offset somewhat by the greater hygroscopicity of the rayon which is responsible for its higher moisture content under comparable conditions both in the tire and out.¹¹ After the first hundred cycles the hysteresis of all cotton and rayon cords are of the same order of magnitude with the possible

¹¹ H. Wakeham, E. Honold, and H. J. Portas, *Ind. Rubber World*, 113, 659 (1946).

exception of the experimental cotton-1 which is somewhat higher than the others over the whole range.

The effect of moisture on the elastic modulus at 25°C as a function of cycle number is depicted in Fig. 7. As the oscillatory loading treatment progresses the elastic moduli of all cords increase at a fairly uniform rate—approximately 0.4×10^{10} dynes per cm^2 for each tenfold increase in cycle number. In Fig. 8 the elastic modulus after 10,000 cycles is plotted as a function of moisture. The cottons exhibit different curves from those for the rayons and Nylon in that the modulus is greatest in the vicinity of 4 to 5 percent moisture. One explanation which might be advanced for this phenomenon is that the swelling by water of the less organized cellulose chains in the inter-micellar regions results in a better stress distribution—as water is added to the dry state the resistance to deformation and the elastic modulus increases. Continued swelling beyond about 5 percent water content, however, increases the

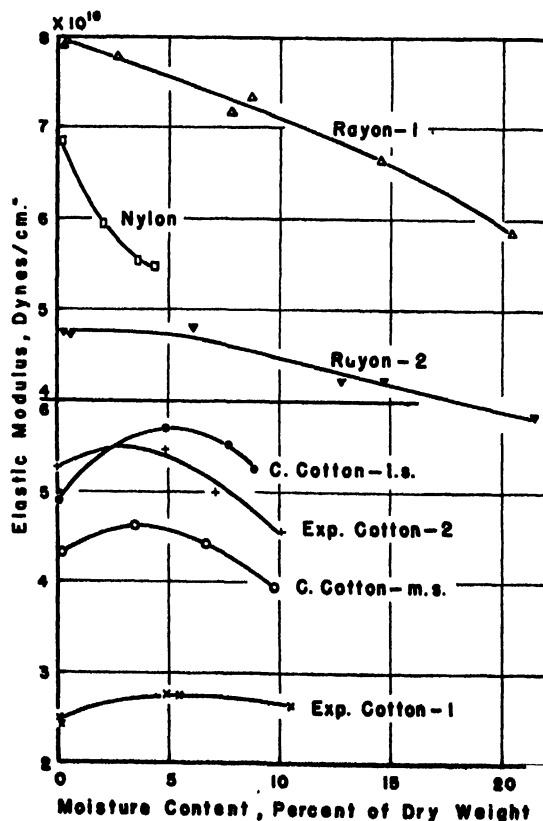


FIG. 8. Elastic modulus vs. moisture after 10,000 cycles at 25°C.

distances between chains to the point where the intermolecular forces decrease and less stress is required to produce unit deformation. This latter effect is observed over the whole moisture range for the Nylon and rayon cords.

A comparison of the elastic moduli of the two rayon cords is of interest inasmuch as the 2200/2 rayon is approximately twice the weight and strength of the 1100/2 cord. The heavier cord has a much lower modulus despite the fact that the modulus calculation takes into account the differences in cross-sectional area of the two cords. As was pointed out earlier, the comparison of these two cords with the same applied tension was, in effect, comparable to running one at two different load ranges. The results shown would seem to indicate that for viscose rayon the elastic modulus for reversible deformation is a function of the load range and increases with increasing load over the range studied.

Typical curves of the elongation as a function of imposed load cycle are plotted in Fig. 9 for dry samples of the various cords studied. It is to be noted that these elongations, which are based on the original length of the cord with a 4-ounce tension, are not accurately reproducible because

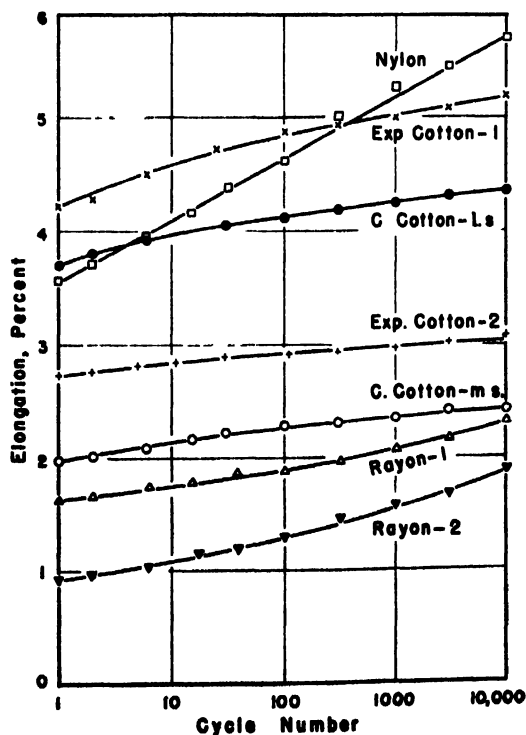


FIG. 9. Elongation vs. cycle number for dry cords at 25°C.

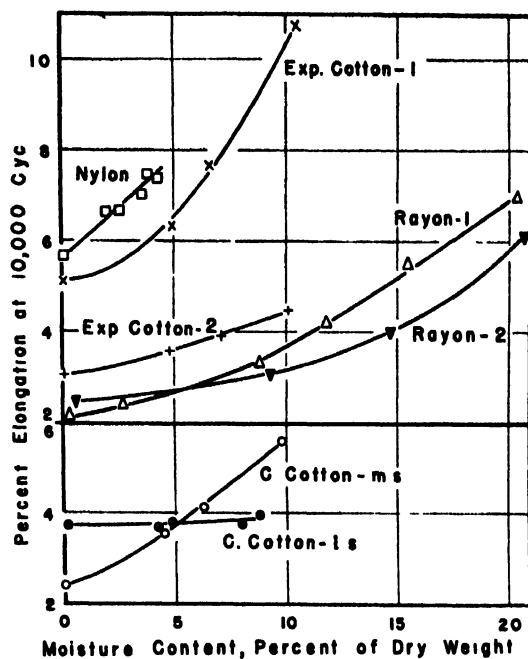


FIG. 10. Total elongation after 10,000 cycles vs. moisture at 25°C.

of some unavoidable variations in handling when the cord is inserted into the oven for testing. The slope of the curve, or the rate of growth, is not appreciably affected by this manipulative error and is a significant characteristic of the cord. Slopes of the elongation curves in Fig. 9 are seen to vary considerably for the different cords.

The total elongation observed during the first 10,000 cycles of cyclic loading is a function of moisture content as shown in Fig. 10. All cords exhibit an increase in elongation with increasing moisture. The cottons with a low stretch (Exp. cotton-2 and C. Cotton 1s.) increase much less than the others. In fact, the low stretch commercial cotton does not increase in elongation at all with moisture addition, simulating its hysteresis behavior mentioned in connection with Fig. 5.

In Fig. 11 are depicted the rates of growth as functions of moisture content. The rayons and high stretch experimental cotton-1 are found to be greatly influenced in their growth rates by moisture. This behavior is in line with the hypothesis that in a fiber with considerable intermicellar space the swelling effect of water decreases the intermolecular forces between the

high polymer chains in these spaces. This allows the material to undergo greater creep and exhibit the progressive growth demonstrated in Figs. 9 and 11. Cotton cords which have been subjected to stretch treatments to reduce their elongations are not only more compact in their cord organizations but also, it is believed, more oriented with respect to the cellulose chains in their intermicellar regions. The intermolecular forces are strong enough to reduce the interference by water molecules. Dillon and Prettyman¹² thus found that tenacity and ultimate stretch in a tensile test of low-stretch cotton cords at room temperature are practically independent of moisture contents, whereas considerable effect was observed for the medium-stretch cotton cord.

It is of interest to compare the curves for Nylon in Figs. 10 and 11. The rate of growth in Fig. 11 is independent of moisture; yet, according to Fig. 10, the total elongation observed after 10,000 cycles increases appreciably with moisture increase. Apparently the initial stretch during the first cycle is a function of moisture, but the

creep factor with continued loadings is not. This is in agreement with Leaderman's¹³ findings on the effect of humidity on creep of Nylon with a static load. The initial deformation observed was considerably greater at the higher humidity than at the lower humidity, but the slopes of the curves at various humidities for deformation *vs.* log time are almost identical.

3. Effect of Temperature

The effect of increasing temperature on the elastic properties of the tire cords herein described is, in general, the same as that of increasing moisture content. The magnitude of the effect due to temperature is, however, somewhat smaller than that of moisture over the ranges commonly encountered.

Curves of hysteresis *vs.* cycle number at temperatures of 25°, 65°, and 145°C are plotted for the dry tire cords in Figs. 12 and 13. Measurements were also made at 105° but the curves for these were omitted for the sake of simplicity in the figures—they were, within experimental error, all between those for 65° and 145°C. It is to be noted that the scale of ordinates in Figs. 12 and 13 is considerably magnified over that shown in Figs. 5 and 6. Except in the Nylon tire cord, the temperature effect on the hysteresis curves is actually a small one, and for the dry cotton cords almost negligible, especially after the first ten thousand cycles.

One characteristic of the rayon cord which is indicated by Fig. 13 deserves particular attention. The 65° hysteresis curves of the rayon cords are lower than those for either 25° or 145°, and are also lower than those for 105°C. In fact, the 105° curves were found to be almost superimposed on those for 25° and, therefore, were not plotted. This result indicates that with increasing temperature on the dry rayon cord the hysteresis first decreases somewhat before rising. Whether or not this phenomenon persists much beyond the ten thousand cycle point is not indicated by the data.

The effects of temperature on the elastic modulus are illustrated in Fig. 14. A small or negligible increase in elastic modulus is shown for the commercial cotton cords in going from 25° to 145°. The rayon modulus, on the other

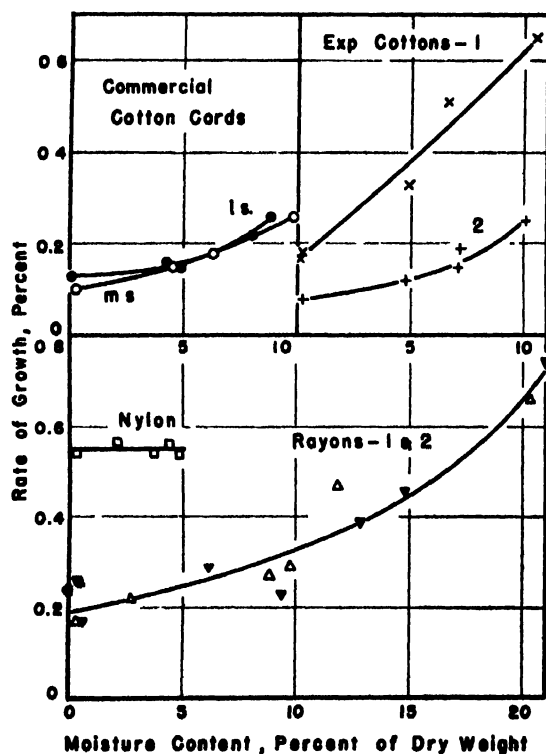


FIG. 11. Rate of growth *vs.* moisture for cords at 25°C.

¹² J. H. Dillon and I. B. Prettyman, *J. App. Phys.* 16, 159 (1945).

¹³ H. Leaderman, *Elastic and Creep Properties of Filamentous Materials and Other High Polymers* (The Textile Foundation, Washington, D. C., 1943) p. 231.

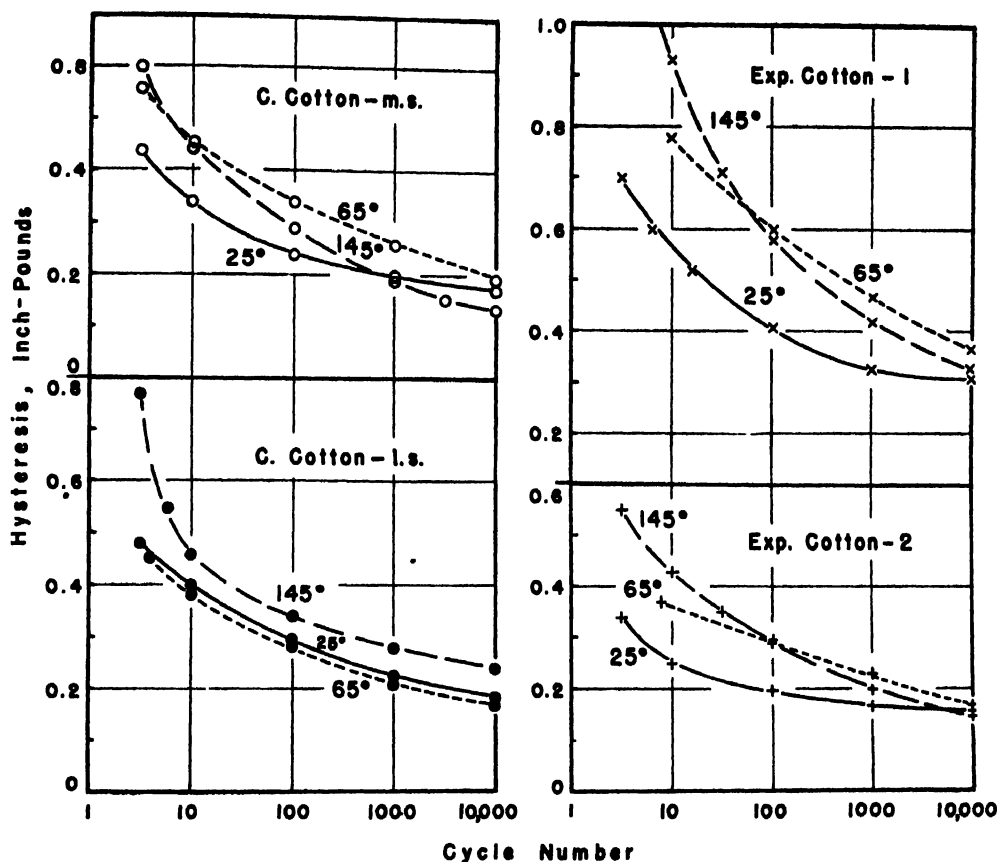


FIG. 12. Hysteresis vs. cycle number at constant temperature for dry cotton cords.

hand, drops considerably over the same temperature range, and the Nylon even more so. This greater reversible elongation which the rayon and Nylon cords show at elevated temperatures may be of some advantage in the tire under road conditions where considerable flexing of the tire is required.

In Fig. 15 is plotted the rate of cord growth, or the elongation per tenfold increase in cycle number, vs. temperature for all seven cords in a dry condition. The cotton cords have low growth rates which increase more or less regularly with temperature; the growths of the rayon and Nylon are high to begin with and increase more rapidly with higher temperatures. These results are in keeping with the known greater growth of rayon and Nylon tires.

4. Combined Effect of Temperature and Moisture Content

After investigating the effects of the moisture content and temperature variables individually,

a thorough survey of the combined effect of these variables did not seem necessary. Since for a given cord increase in either temperature or moisture generally produced similar changes in the elastic properties studied, it seemed logical to expect that increasing both factors simultaneously would merely result in a greater change.

The results shown in Fig. 16 on the medium stretch commercial cotton and the rayon-1 cords justify such an expectation. For the cotton cord, going from 25° and dry cord either to 25° and 3 percent moisture or to 100° and dry cord produces almost identical changes in elastic properties; similarly for the rayon a change either to 25° and 6 percent or to 100° and dry cord results in the same effect. Combining the temperature and moisture increases produces the new curves shown.

In terms of the intermolecular forces between the high polymer chains in the intermicellar

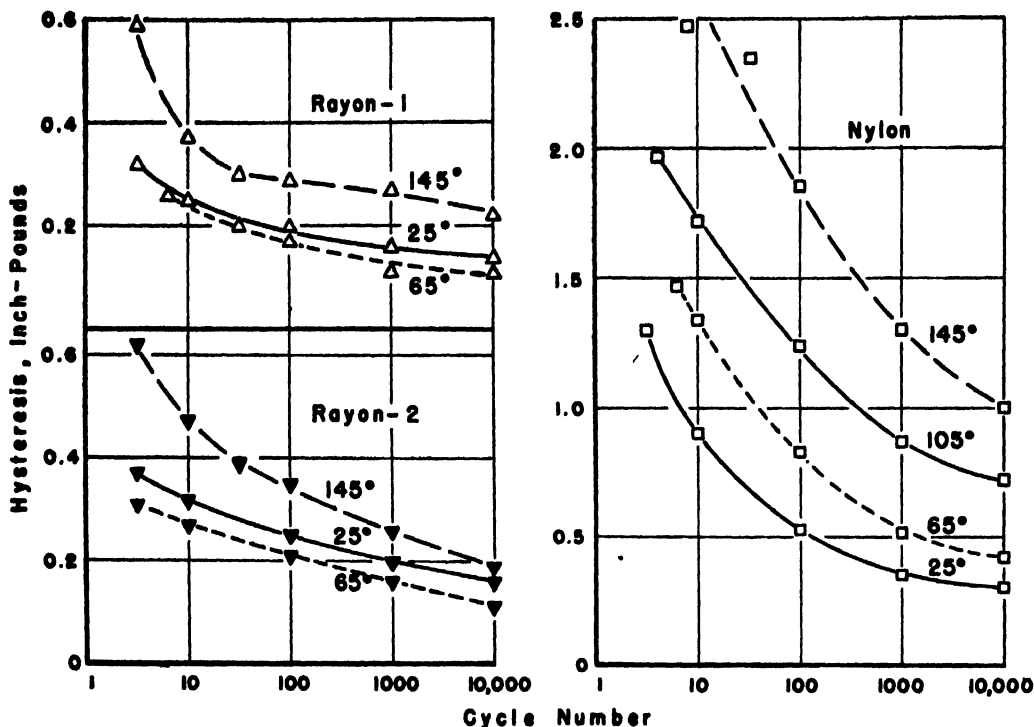


FIG. 13. Hysteresis vs. cycle number at constant temperature for dry rayon and Nylon cords.

spaces of the cellulose fiber, the similarity in elastic behavior produced by high temperature and high moisture content is quite reasonable. The explanation has already been advanced that the addition of water molecules to the regions of unoriented chains decreases the intermolecular forces existing between the chains. An increase in temperature would likewise be expected to de-

crease the same forces and result in greater elongation and growth, and lower elastic modulus.

5. Effect of Load

The results so far reported have all been for cyclic loads over the same range of 1-4 pounds regardless of the cord being tested. It seemed reasonable to expect that the elastic properties of a cord should vary somewhat with the load range chosen for the cycle. An experimental exploration of this effect was possible with the present apparatus using the rayon-1 cord. The load range was reduced to approximately 1.6 pounds (instead of 3 pounds) and runs were made at five mean loads varying between 1.3 and 4.6 pounds. Elastic properties after 1000 cycles vs. the mean load are plotted in Fig. 17. It is to be noted from this figure that at low mean loads the hysteresis of the cord rises appreciably to higher values. This result is of practical interest since the average tensions on individual cords in the tire are thought to be fairly low.

In view of this effect of load on the elastic properties of a tire cord, it appeared worth while to make a comparison of these properties when

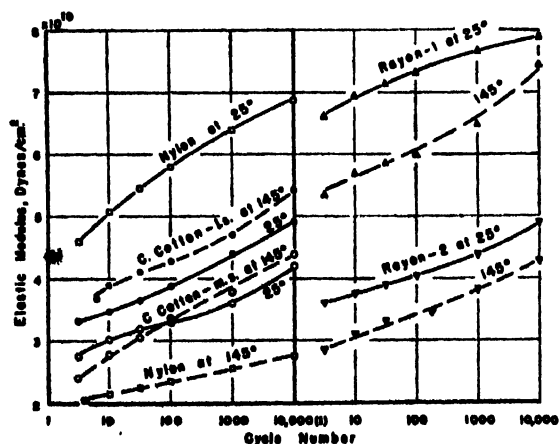


FIG. 14. Elastic modulus vs. cycle number at constant temperature.

the load cycle is on a comparable basis—for example, over the same range in grams per grex for each cord. The range 0.1 to 0.4 gram per grex was chosen and runs were made on all cords in the dry condition at 25°C. The values obtained after 1000 cycles, shown in Table II, demonstrate that the properties of these cords when compared on this basis vary widely not only when cords of different types are considered but even when cords of the same group, i.e., cotton, are compared.

6. Effect of Cyclic Loading on the Stress-Strain Curves

In commercial testing much importance as a quality index is attached to the tenacity and elongation of a tire cord as measured by a conventional testing machine. Cords to which the cyclic loading treatment had been applied for 10,000 cycles were tested for their stress-strain curves in such a testing machine and the curves compared with similar ones for untreated samples of the same cord. These are shown in Fig. 18. Since it was necessary to condition all cords to comparable moisture levels before testing, a tension of 0.5 pound was applied to the cord during the conditioning period in order to prevent undue relaxation during this time. Thus, in Fig. 18 there are shown for each cord the stress-strain curves for (1) the conditioned cord directly from the spool; (2) the conditioned cord to which a tension of 0.5 pound had been applied during conditioning; and (3) the conditioned cord which had been subjected to cyclic loading and to which a 0.5-pound load had been applied during conditioning. The resulting curves demonstrate that the cyclic loading process shifts the stress-strain curves to lower strain or elongation values.

DISCUSSION

In the foregoing experiments the hysteresis, elastic modulus, elongation, and rate of growth of several tire cords have been compared under cyclic loads similar to those applied in certain fatigue tests. It has been found that these cord properties vary with cycle number, moisture content, temperature, and load. It is unfortunate that the apparatus used in these experiments does not also permit investigation at higher rates

of cyclic loading, as this is very probably an additional factor influencing cord properties. Such a study must be left for future investigations.

The most significant of the variables investigated appears to be the moisture content of the tire cord. The properties of all types of cords measured are affected to a greater or less degree by this factor, probably because the polymeric structure of the fibers is influenced by the added water. Some of the moisture may be thought of as entering into the intermicellar spaces and finding places between the cellulose chains in these regions. There the intermolecular forces of the water molecules are revealed as a viscous drag, increasing the hysteresis of the fiber. The water molecules, furthermore, may decrease somewhat the cohesion between chains permitting greater elastic stretch when a load is applied. Though such an explanation may not be the correct one, it emphasizes the fundamental influence which moisture exerts on tire cord properties and the extent to which moisture should be considered in problems of fabric design and utilization.

It seems reasonable to presume that the fatigue process which the cord undergoes in a tire, if at

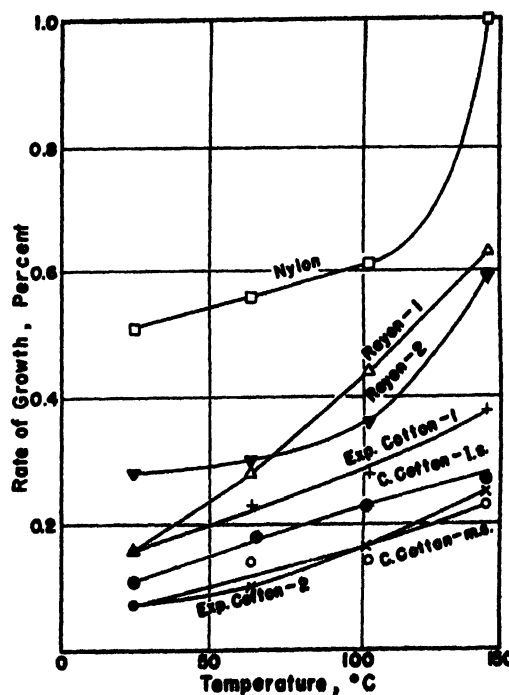


FIG. 15. Rate of growth vs. temperature for dry cords.

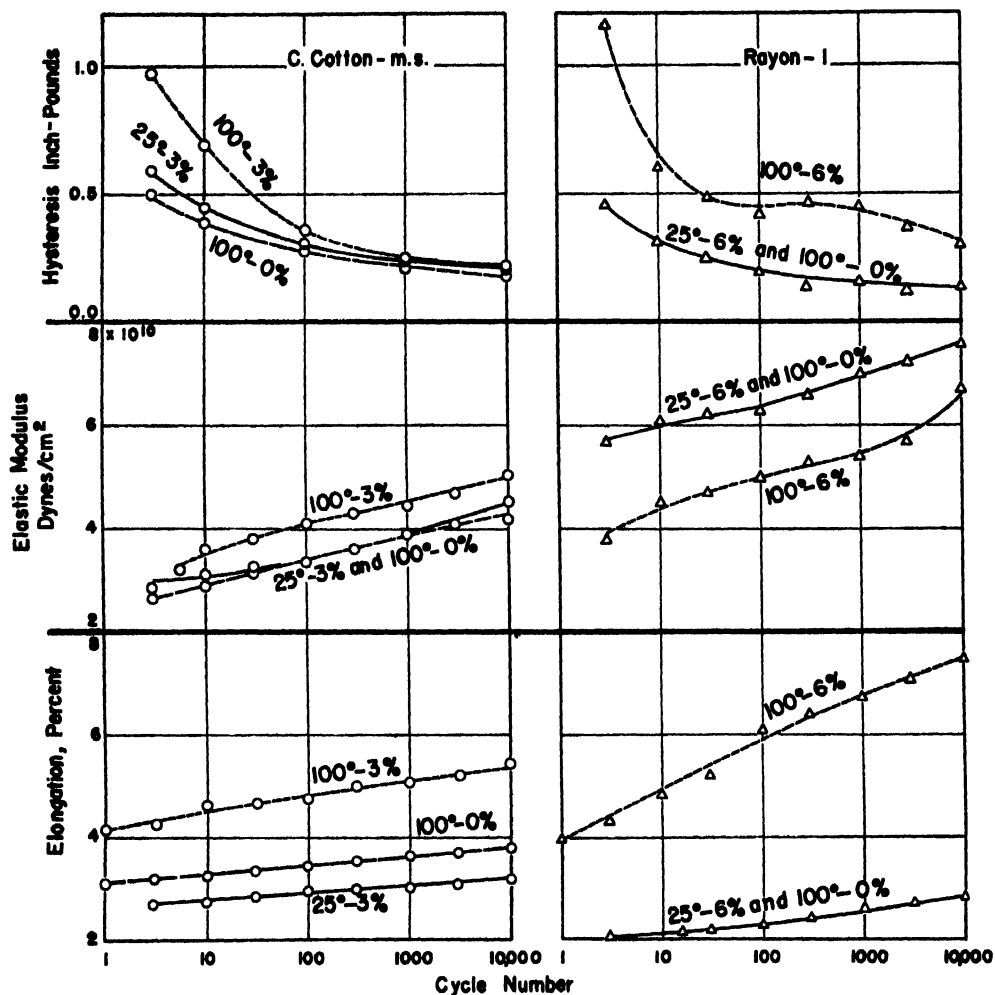


FIG. 16. Effect of temperature and moisture combined. Hysteresis, elastic modulus, and elongation vs. cycle number for a cotton and a rayon tire cord.

all similar to that involved in these experiments, is slower than that followed in an accelerated laboratory fatigue test. Certainly the cord in the tire must go through many more load cycles than the few hundred thousand required to produce failure in the laboratory. This difference is probably associated with the fact the cord in the tire is surrounded by rubber. The rubber restricts the motion and deterioration of mechanical properties; and this restricting effect is absent in the cord tested in the laboratory.

How significant hysteresis and elastic modulus of tire cords are in tire performance is, therefore, a problem difficult to evaluate. It seems certain that hysteresis energy of the cord contributes to heat build-up in the tire. The authors in another investigation, for example, have demonstrated

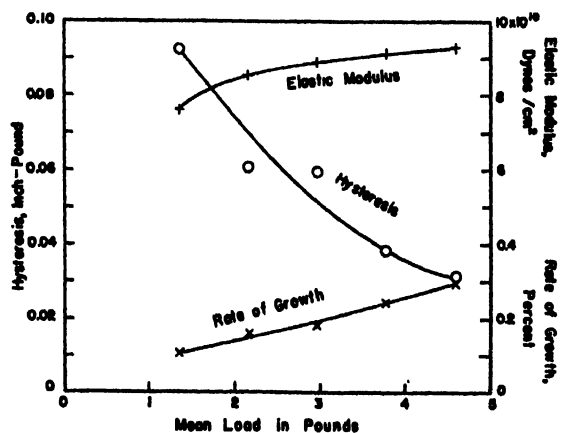


FIG. 17. Effect of load. Hysteresis, elastic modulus, and rate of growth vs. mean load for dry rayon-1 at 25°C after 1000 cycles had been run. A load range of 1.6 pounds was maintained for each mean load plotted.

TABLE II. Comparison of elastic properties of tire cords after 1000 cycles
(25°C, dry cords, 0.1 to 0.4 g/grex).

Cord	Grex (dry)	Elastic modulus dynes/cm ²	Growth rate at 0.2 g/grex	Hysteresis Inch-lb. per 100 inches	Calories per gram
Nylon	1986	6.00×10^{10}	0.247%	0.098	5.25×10^{-4}
Rayon-1	2444	7.66	0.105	0.072	3.10
Rayon-2	5366	4.60	0.215	0.258	5.10
Com. Cotton (m.s.)	4342	3.56	0.073	0.207	5.07
Com. Cotton (l.s.)	5064	5.16	0.130	0.278	5.83
Exp. Cotton-1	4716	2.07	0.170	0.356	8.01
Exp. Cotton-2	4446	4.37	0.086	0.169	4.05

that the torsional hysteresis of tire sections containing cord are higher than the same sections in which the cord structure has been destroyed.¹⁴ But tire heat build-up is determined by so many additional factors involved in tire structure and rubber composition that the part contributed by the cord can only be estimated. A 7.00-20 truck tire, weighing approximately 65 pounds, contains about 12 pounds of cord. Assuming that the average stress cycle of the cord is such that the hysteresis energy loss is approximately 5×10^{-3} cal. per gram of cord (Table II), one may estimate that the fabric produces around 27 cal. of heat per revolution of the tire. At 500 r.p.m., or approximately 50 miles per hour, the cord is generating about 13,500 cal. of heat each minute. Since such a tire has a heat capacity of about 10,000 cal. per degree C, the heat from the cord is sufficient to produce, under adiabatic conditions, a temperature rise of about 1.35°C per minute. This estimate indicates without much question that the elastic hysteresis of the cord fabric makes an appreciable contribution to the high temperatures observed in tires in service.

One might deduce from a consideration of tire design and performance that a low elastic modulus in the tire cord would be desirable. Greater reversible stretch in the cord should produce a tire more homogeneous in elastic properties, with fewer ply and tread separations and increased bruise resistance. Here again, however, it is only possible to surmise a correlation between laboratory testing and tire performance without making tests with actual tires.

¹⁴ H. Wakeham and E. Honold, Ind. Rubber World 113, 377 (1945).

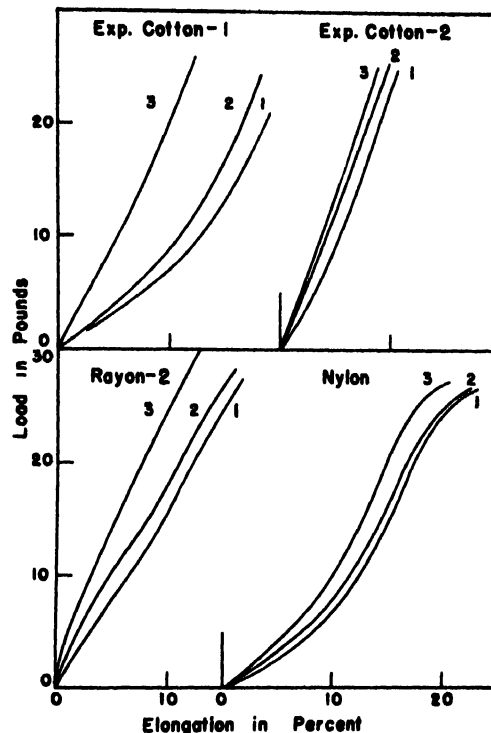


FIG. 18. Stress-strain curves for (1) moisture conditioned cord, (2) cord conditioned with a tension of 0.5 pound, and (3) cord subjected to cyclic loading for 10,000 cycles and conditioned with a tension of 0.5 pound.

ACKNOWLEDGMENT

The authors are grateful to many members of the Southern Regional Research Laboratory staff for their interest and cooperation in this work, and especially to Jacquelyn Souderes and Gene Hassinger for their assistance with many of the calculations.

Letters to the Editor.

On the "Power Function"

FIDEL ALSINA FUERTES

Agrupación Estudiantes de Física, Instituto de Física, La Plata, Argentina
December 12, 1945

IN a recent note, D. A. Wells¹ sets up a "power function" through which the generalized Lagrangian forces are determined for a wide variety of forces acting on a holonomic system, by the relation

$$Q_k = \partial P / \partial \dot{q}_k$$

(P = power function; $\dot{q}_k$ = time derivative of the generalized coordinate q_k).

When the forces are proportional to the speed of the particles acted upon, P reduces to the dissipative function considered by the late Lord Rayleigh.² The purpose of this letter is to show to what extent the P -function widens the field of allowable forces.

Let us consider, for the present, a single particle. For such a simple system, the P -function reads

$$P = \int \sum_{i=1}^3 F_i \dot{x}_i \quad (1)$$

(F_i = cartesian component along the i -axis of the force acting on the particle; $\dot{x}_i$ = time-derivative of the i -Cartesian coordinate).

Integration limits, not shown in (1), must be chosen such that

$$\partial P / \partial \dot{x}_i = F_i. \quad (2)$$

P is, according to (1), a function of both upper and lower limits, and perhaps of the integration path. One of the limits is taken constant to obtain (2), and the independence of the integration path is assured when the "conditions of integrability"

$$\partial F_i / \partial \dot{x}_j = \partial F_j / \partial \dot{x}_i \quad (3)$$

are fulfilled. When $F = av^n$, Eq. (3) holds (Wells). When F_i are proportional to $\dot{x}_i$ (Rayleigh), Eq. (3) holds automatically, since both sides are zero.

Let it be now F a force depending on the speed v of the particle in an arbitrary way

$$F = f(v); \quad (4)$$

$f(v)$ may contain, further, coordinates and time. If the direction of F coincides with that of v , it can be easily shown that Eq. (3) holds for any $f(v)$. We have

$$\Leftrightarrow F_i = f(v) \frac{dx_i}{ds} = f(v) \frac{\dot{x}_i}{v} \quad (5)$$

and, taking into account $v^2 = \sum \dot{x}_i^2$,

$$\frac{\partial F_i}{\partial \dot{x}_j} = \frac{d}{dv} \left(\frac{f(v)}{v} \right) \frac{\dot{x}_j \dot{x}_i}{v} \quad (6)$$

a relation asserting the validity of (3).

Thus, for single-particle systems, P -functions can be obtained under much more general conditions than allowed by "dissipative functions." In spite of this, it is interesting to remark that anisotropic forces cannot enter in consideration, lest (6) ceases to be symmetrical in i, j . The only

case of allowable anisotropy appears when the forces are proportional to the speed, i.e., Rayleigh's case.³

Let us proceed now to systems formed by p particles. In order to set up the P -function for the system, we first obtain the P -function for each of the p particles and then add up

$$P = \sum_{\nu=1}^p P_{\nu} \quad (\nu = 1, 2, \dots, p). \quad (7)$$

The existence of P is restricted by same conditions as above, namely: *forces should be isotropic, directed along the tangents to the trajectories of the particles acted upon, and depending on the speeds, coordinates, and time in an arbitrary way.*

When considering many-particle systems, one is tempted to put simply a new sum sign in (1). By doing so, Eq. (3) would appear

$$\frac{\partial F_{\nu i}}{\partial \dot{x}_{\mu j}} = \frac{\partial F_{\mu j}}{\partial \dot{x}_{\nu i}}$$

and the functions (4) should be taken independent of the individual particle in order to satisfy (6). This is a drastic restriction, as it cuts off even simple applications like Example 1 in Wells' article. Besides, it is an unnecessary limitation, since definition (7) does not differ from Wells'. In fact, the expression for P remains

$$P = \sum_{\nu=1}^p \int \sum_{i=1}^3 F_{\nu i} \dot{x}_{\nu i} \quad (8)$$

with the integrability conditions

$$\frac{\partial F_{\nu i}}{\partial \dot{x}_{\nu j}} = \frac{\partial F_{\nu j}}{\partial \dot{x}_{\nu i}} \quad (9)$$

It is easy now to find a simpler expression for P . Let us put, according to (4)

$$F_{\nu} = f_{\nu}(V_{\nu});$$

the i -component of the force acting on the ν -particle reads

$$F_{\nu i} = f_{\nu}(v_{\nu}) \frac{\dot{x}_{\nu i}}{v_{\nu}} \quad (10)$$

and, using (8)

$$P = \sum_{\nu=1}^p \int f_{\nu}(v_{\nu}) \frac{1}{v_{\nu}} \left[\sum_{i=1}^3 \dot{x}_{\nu i} d\dot{x}_{\nu i} \right]. \quad (11)$$

The bracket can be written

$$\sum_{i=1}^3 \dot{x}_{\nu i} d\dot{x}_{\nu i} = \frac{1}{2} d \sum_{i=1}^3 \dot{x}_{\nu i}^2 = \frac{1}{2} dv_{\nu}^2 = v_{\nu} dv_{\nu} \quad (12)$$

and finally

$$P = \sum_{\nu=1}^p \int f_{\nu}(v_{\nu}) dv_{\nu}. \quad (13)$$

This expression is obviously independent of the frame of coordinates. The rest of the equations can easily be rewritten in generalized coordinates, if desired.

Since the power-function has been specified, and used, only in relation to $\dot{x}_i$'s (or the $\dot{q}_i$'s), it is evident that it is determined for each system except for an additive arbitrary function of coordinates and/or time

$$P = \sum_{\nu=1}^p \int f_{\nu}(v_{\nu}) dv_{\nu} + \varphi(x_i, t).$$

¹ D. A. Wells, J. App. Phys. 16, 535 (1945).

² Lord Rayleigh, in *Collected Papers* (1899), Vol. 1, p. 176.

³ E. T. Whittaker, *Analytical Dynamics* (1937), p. 231.

Journal of Applied Physics

Volume 17, Number 9

September, 1946

Internal Friction and Plastic Extension of Zinc Single Crystals

THOMAS A. READ*

Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania

AND

E. P. T. TYNDALL

Department of Physics, State University of Iowa, Iowa City, Iowa

(Received April 29, 1946)

This paper presents data on the internal friction of four single crystals of zinc while oscillating longitudinally, and a description of various slow speed tension tests on a fifth crystal within and beyond the elastic limit. In the first named measurements the behavior of the crystals bears little resemblance to that of crystals of zinc of greater purity prepared by another method, which has previously been reported (see reference 1). The most outstanding feature is that the decrement, although higher at the lowest stress amplitude than for the previous ones, shows very little rise with increasing stress amplitude, even up to stresses far beyond the statically-determined elastic limit. The difference in behavior seems to be caused by the difference in purity of the zincs. An optically mosaic

structure, such as that described by Schilling (see reference 4), does not appear to be responsible. In the slow speed tension tests the zinc crystal, after a period of self-annealing at room temperature, has a Hooke's law region up to about 70 g/mm² (R.S.S. = 28 g/mm²) and at greater stress shows transient and steady creep. Under some circumstances the transient creep is in two parts, a slow starting creep followed by much more rapid creep. The crystal, when measured following a permanent strain, shows creep at loads well below the previous elastic limit and has a large amount of hysteresis. Self-annealing occurs during rest periods and the initial elastic limit is gradually regained while the hysteresis disappears.

INTRODUCTION

THE only previous work on the internal friction of zinc single crystals is that described by one of the authors in a previous paper.¹ The specimens used in this earlier investigation were grown from spectrographically pure zinc (99.999 percent Zn) by the Bridgman method in Pyrex glass tubes coated with graphite. The damping behavior of these crystals was very complicated, but may be briefly summarized by saying that

after a one- or two-day anneal at room temperature their internal friction was very low (decrement about 1×10^{-8}) at small amplitudes of vibration, but increased with increasing amplitude. This increase was very rapid for crystals whose hexagonal axis made an angle of from 15 to 75 degrees with the axis of the rod. For example, the decrement of a crystal for which this angle was 74.5° was 1.7×10^{-8} at a stress amplitude of .15 g/mm², and 144×10^{-8} at a stress amplitude of 22.2 g/mm². Both these figures for the stress refer to the values at displacement nodes. The corresponding values of the strain amplitude are 1.2×10^{-8} and 176×10^{-8} .

* Westinghouse Research Fellow. Now at the Ordnance Laboratory, Frankford Arsenal, Philadelphia, Pennsylvania.

¹ T. A. Read, *Phys. Rev.* **58**, 371 (1940).

The objects of the investigation described in the present paper were to compare the internal friction of zinc single crystals grown by another method with that of the Bridgman crystals, and to study the relation between internal friction and behavior in slow tension tests.

SPECIMEN MATERIAL

The zinc crystals used in the present investigation were grown from "Bunker Hill" zinc, a commercial zinc whose total impurity content

TABLE I. Description of crystals.

Designation	Orientation	R.S.S. factor	Remarks
2	15°	0.25	O.M.
3	72°	0.29	G.C. Pb, Sn
4	70°	0.32	G.C.
5	81°	0.17	O.M. Pb, Sn
6	63°	0.41	G.C.

O.M.: Optically mosaic.
G.C.: Good crystal.
Pb, Sn: Lead and tin impurities.

is about 0.01 percent. It is to be noted that this purity is somewhat less than that of the spectrographically pure zinc used in the earlier investigation referred to above. At the conclusion of the work a spectrographic analysis was made on four of the crystals and on a small ingot cast from the original material. It revealed that crystals Nos. 2 and 4 were identical with the original material but that Nos. 3 and 5 contained excess amounts of lead and tin,² which had apparently been introduced accidentally by the use of a "seed" crystal contaminated with these metals. The effect of the difference in purity between the two pairs of crystals was very marked as may be seen below.

These crystals were grown by the modified Kapitza method described by Cinnamon.³ The molds were made of well-baked transite with trapezoidal grooves, 0.25 cm² in cross section for crystals Nos. 2-5 and 0.5 cm² for crystal No. 6. The surface of specimens grown this way is somewhat rougher than that of crystals grown in glass molds. This is caused by the wrinkling of a very thin surface film of oxide when the

metal contracts in freezing. The method is believed, however, to be more likely to produce more nearly perfect crystals because of lack of constraint on the growing crystal. Two of the crystals were, however, purposely grown of the "optical mosaic" type.^{4,5}

The specimens are described in Table I. The orientation is given as the angle between the hexagonal axis and the specimen length. The third column gives the factor by which a longitudinal stress must be multiplied to obtain the resolved shear stress in the basal (glide) plane. Further resolution into the glide direction would require an additional factor which must, however, always lie between 1.0 and 0.866. Any deviation of this from 1.0 has been neglected in the following.

The decision as to the presence of an optically mosaic structure was made on the basis of examination of the portions of the crystal adjacent to the specimen, and of the specimen itself after the completion of the measurements.

After growth the first four crystals were cut to approximately the proper length for one half-wave of longitudinal vibration with a jeweler's hack saw run automatically and so arranged that the blade pressed very lightly against the crystal. They were then packed in Vaseline in brass tubes which were in turn packed in cotton in a sturdy box. This box was mailed from Iowa City to East Pittsburgh. The authors feel that the care taken in this handling of the specimens sufficed to insure that their internal friction was not affected thereby. The results of the measurement are consistent with this view with one exception, as noted below. In East Pittsburgh the crystals were cut to exact length with a jeweler's saw after being waxed in two V-blocks with beeswax and rosin. One end of each crystal was ground flat by holding the crystal lightly in a V-block and grinding it on a cast iron surface plate. This same method of preparation was employed in the previous work.¹ Crystal No. 6 was mounted in the measuring apparatus as grown. The crystal was 15 cm long and measurements were made on a central 9 cm length.

¹ In crystals Nos. 2 and 4 and the ingot Sn was either not detected or less than 0.001 percent, while Pb was either not detected or less than 0.002 percent. In Nos. 3 and 5 Pb was about 0.01 percent, and Sn 0.1 percent (No. 3) and 0.03 percent (No. 5).

² C. A. Cinnamon, *Rev. Sci. Inst.* 5, 187 (1934).

⁴ H. K. Schilling, *Physics* 6, 111 (1935).

⁵ Electrical and thermal resistivities of optically mosaic crystals differ in an erratic way from those lacking this structure. W. J. Poppy, *Phys. Rev.* 46, 815 (1934) and C. A. Cinnamon, *Phys. Rev.* 46, 215 (1934).

INTERNAL FRICTION

The internal friction of these crystals was measured for longitudinal vibration at a frequency of 33.5 kilocycles per second by the composite piezoelectric oscillator method. The apparatus used was similar to that previously described.¹ The dependence of the internal friction of the specimen crystals on the amplitude of vibration was obtained by measuring the impedance of the composite piezoelectric oscillator as a function of the voltage impressed on electrodes attached to the quartz and of frequency in the neighborhood of the resonant frequency. It can be readily shown that the stress amplitude σ at a displacement node in the specimen is given by the formula

$$\sigma = \frac{2E\delta(10)^{1/2}K}{R_0V} \times 10^3 \text{ dynes/cm}^2, \quad (1)$$

where E = Young's modulus in dynes/cm² of the specimen in the direction of its length.

δ = r.m.s. voltage applied to electrodes.

K = proportionality constant between equivalent inductance of composite piezoelectric oscillator and its mass in henrys per gram.

R_0 = resistance at resonance in ohms for given value of δ .
 V = velocity, in cm/sec., of sound in the specimen in the direction of its length.

The formula (1) was given in reference 1 in a somewhat less convenient form.

For all of the measurements the specimens

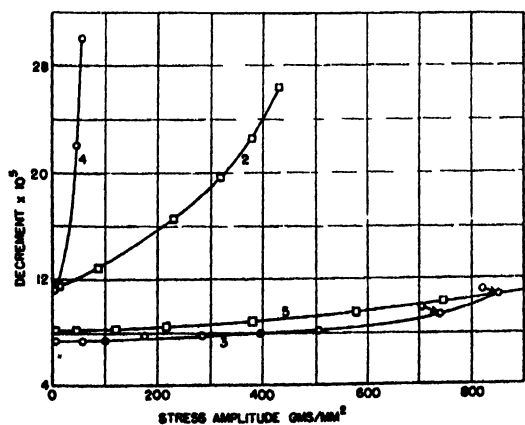


FIG. 1. The internal friction of the four crystals investigated as a function of the stress amplitude at a displacement node. The arrows indicate a change with time at constant driving voltage. The crystal number is indicated on each curve.

were suspended in a vacuum and maintained at a temperature of 23°C.

A convenient measure of internal friction is the decrement, defined as the ratio of the energy dissipated in the specimen in a cycle of vibration

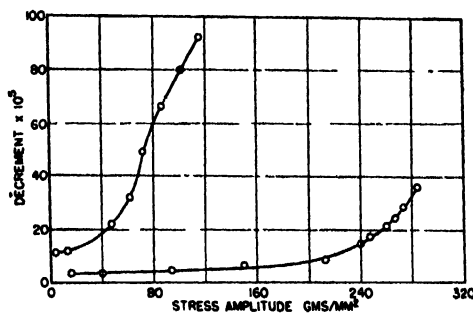


FIG. 2. The lower curve gives the internal friction of crystal number 4 after a one hour anneal at 330°C. The upper curve, repeated from Fig. 1, shows the original internal friction.

to twice the total energy stored in the vibration of the specimen.

The dependence of the internal friction of each of the four specimens on the amplitude of vibration is given in Fig. 1. These data, obtained without any annealing or other treatment except as described above, are remarkable for five reasons. First, the recovery period of a day or two after mounting, during which the previously measured crystals¹ exhibited a decrease in internal friction by a factor of from 10 to 100, was completely absent for these crystals, although exactly the same technique of affixing the specimens to the quartz rod was used. Second, the internal friction of the two crystals of lesser purity is only slightly greater at stress amplitudes of several hundred g/mm² than at very small amplitudes of vibration. Third, there is no marked difference between the behavior of the crystals of the same purity which have an optically mosaic structure and those which do not (compare Nos. 2 and 4 and Nos. 3 and 5). Fourth, these crystals have considerably higher internal friction at small amplitudes of vibration than did the zinc crystals described in reference 1. Fifth, the change of the internal friction with time, at constant driving force, which occurs at the largest amplitudes is a *decrease*. This change is indicated by the arrows.

Data were obtained on three methods of

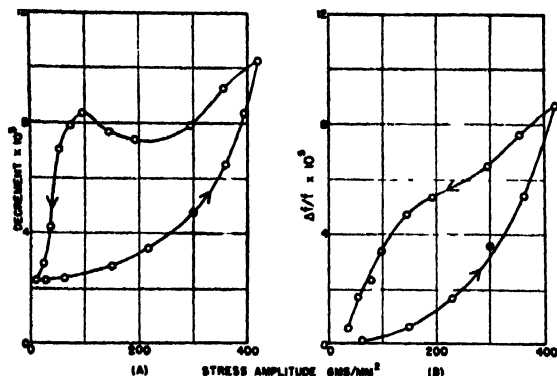


FIG. 3. (a) The internal friction of crystal number 3 as a function both of increasing and decreasing amplitude of vibration. The sequence of the measurements is indicated by the arrows. (b) The changes in the natural frequency of longitudinal vibration of crystal specimen number 3 which are observed simultaneously with the decrement values plotted in Fig. 3(a).

changing the internal friction of these crystals, *viz.*, etching, annealing, and cold-working.

Etching one of these crystals does not appreciably affect the dependence of internal friction on amplitude, but does produce a lowering of the curve as a whole. For example, a three-minute immersion of crystal No. 3 in 30 percent HCl decreased the small-amplitude decrement from 7.6×10^{-3} to 4.0×10^{-3} . This treatment decreased the weight of this crystal by 0.05 percent. Subsequent electrolytic etching of this crystal lowered the small-amplitude decrement further to 2.3×10^{-3} . The corresponding loss in weight was 3.8 percent. It may be concluded that the greater part of the original internal friction of this crystal at small amplitudes was a surface effect rather than a property of the material.

As shown in Fig. 1 (and in Fig. 2, where a more extensive curve is plotted), the internal friction of crystal number 4 was considerably higher and rose more rapidly with increasing amplitude than for the other crystals. For this reason this crystal was selected for an investigation of the effect of a heat treatment; it was suspected that this crystal may have been strained by the handling described above. The change in the internal friction which was produced by an anneal of one hour at 330°C is shown in Fig. 2, with the lower curve depicting the results obtained after the anneal. Of the four crystals described in this paper, this particular crystal exhibited the greatest similarity in proper-

ties to the crystals reported on in reference 1. For example, the internal friction of this crystal always increased with time when any change was observed at constant amplitude of vibration.

Cold-working raises the internal friction of these zinc crystals, except in certain cases in which the cold-work is introduced by the internal friction measurement itself. This exception has been mentioned in connection with Fig. 1. The more typical case, in which the cold-working associated with the internal friction measurement raises the internal friction is exemplified by Fig. 3 (a). In this figure is plotted a sequence of measurements on crystal No. 3, beginning at a very small amplitude of vibration, proceeding in steps to a stress amplitude of 420 g/mm^2 , and then decreasing in steps to the initial amplitude of vibration. This curve is similar in shape to one given in reference 1, except that here the internal friction actually rises to a maximum as the amplitude of vibration is decreased. A period of only twenty minutes elapsed during the measurements. After approximately 24 hours the lower branch of the curve may be reproduced. The internal friction of the crystal shows a more rapid rise than that shown in Fig. 1 because shortly before making the measurements for Fig. 3 this crystal was cold-worked statically in the manner described below.

As is described at greater length in reference 1, there is in general associated with the change in the decrement produced by a change of the amplitude of vibration a change in the natural frequency of longitudinal vibration of the specimen rod. This change in the resonant frequency is, in many cases, related in a simple way to the change in internal friction. In the instances cited there a plot of successive values of the change in natural frequency against the corresponding values of the decrement for a series of amplitudes of vibration produced a straight line. That this simple relation does not exist for crystal number 3 under the present conditions of measurement is demonstrated by a comparison of Fig. 3 (a) with Fig. 3 (b) which gives the changes (decreases) in the natural frequency of this crystal which were observed simultaneously with the decrement changes plotted in Fig. 3 (a). It is, however, important to note that a maximum slope of the frequency change curve given in

Fig. 3 (b) occurs at the same amplitude of vibration as the maximum of the internal friction curve of Fig. 3 (a).

The effect on the internal friction of the cold-work introduced by a static load is shown by the difference between the lower and upper curves of Fig. 4, which were taken, respectively, immediately before and after subjecting crystal No. 3 to a compressive stress of 1.8 g/mm² for one minute. The greater part of this increase in internal friction disappeared in the course of two or three days. These data suggest that a previously applied static load which acts for an appreciable time is more effective in causing the decrement to show the rapid rise with increasing stress amplitude (i.e., the specimen acts less elastically) than is a much greater stress which is rapidly alternated. The slow speed tension tests described below seem to afford some clue to this behavior.

SLOW SPEED TENSION TESTS

Various slow speed tension tests were made on crystal No. 6 with the apparatus which has been described briefly elsewhere.⁶ The tests consisted in observing the elongation under load for various conditions of the crystal, which were determined by the nature of the preceding runs, periods of rest between runs, etc. Creep curves under steady loads of various values were also obtained. Representative experiments are described in detail below.

The first run taken on the crystal immediately after mounting it in the apparatus showed slight hysteresis and a lower elastic limit⁷ than was

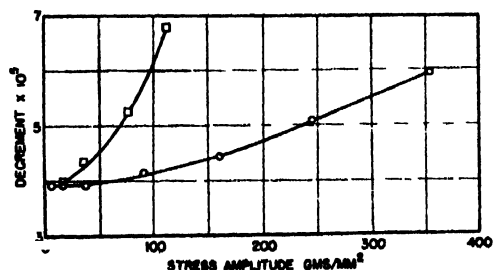


FIG. 4. The increase in the internal friction of specimen number 3 produced by a static compressive stress of 1.8 g/mm² applied for one minute. Note that this static stress is much smaller than the stress amplitude in the internal friction measurements.

⁶ I. H. Swift and E. P. T. Tyndall, *Phys. Rev.* 61, 359 (1942).

⁷ At a stress of 43 g/mm².

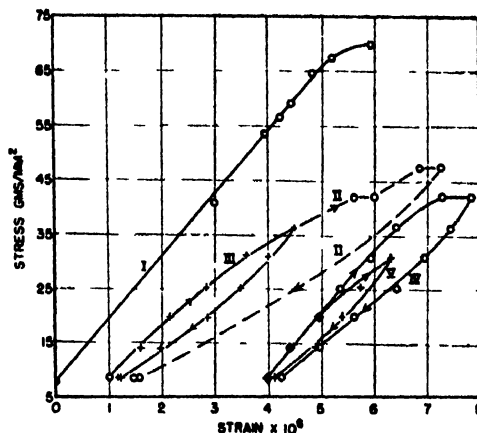


FIG. 5. Crystal No. 6: curve I, in the elastic state. The last load produced a permanent set of $\Delta l/l = 2 \times 10^{-4}$ (not shown on curve). Curves II, III, crystal in inelastic state. Curve IV, after two days rest. Curve V, hysteresis loop produced by small plastic strain at end of curve IV.

expected. A slight permanent set ($\Delta l/l = 1.5 \times 10^{-6}$) was given the crystal at this time. It was therefore allowed to rest under the minimum load (equivalent to a stress of about 7 g/mm²) for two weeks. It was then found to be in an "elastic state" (see later) and the Young's modulus was carefully measured, using the photoelectric multiplying device⁶ and, for safety, with the maximum load kept below the previously found elastic limit. After a total period of 42 days in which the crystal had either rested or been subjected to only small loads, the load was gradually increased. The data obtained are shown in curve I, Fig. 5. It will be seen that the proportional region goes to a stress of 67.2 g/mm². The next load (stress = 69.9 g/mm²) caused creep to an "off-scale" deflection and the specimen was hurriedly unloaded. The creep, however, was observed to occur rather slowly at first and then with increasing rapidity. In fact the point plotted at 69.9 g/mm² corresponds to the scale reading a second or two after completing the loading.⁸ This phenomenon of slow creep followed by much more rapid creep was observed many times in subsequent experiments. The duration of the slow creep is far too long to be accounted for by the inertia of the load or of the moving parts of the observing apparatus. The permanent set left after unloading was

⁸ The loading is done by running mercury into a container hung from the lower end of the crystal. About 20 seconds elapsed between the 67.7 and 69.9 g/mm² readings.

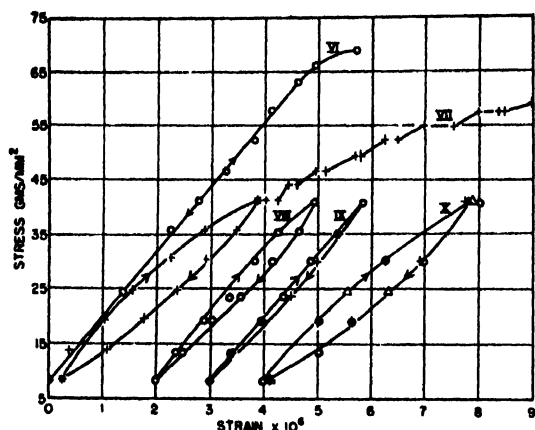


FIG. 6. Crystal No. 6: curve VI, crystal in elastic state after ten days rest. Permanent set at end of curve VI, $\Delta l/l = 1.7 \times 10^{-4}$ (not shown). Curve VII hysteresis loop and creep, immediately following curve VI. Curves VIII and IX, one and three days after curve VII. Curve X, typical hysteresis loops of crystal in inelastic state.

$\Delta l/l = 2 \times 10^{-4}$. The data shown in curve II were taken next.⁹ This curve shows creep at two values of stress (41.4 and 46.9 g/mm²) which were previously within the elastic or proportional region. The time interval between the two points plotted at 41.4 g/mm² is 3 minutes; at 46.9, it is 2 minutes. On unloading there is a slight recovery in 1 minute as shown by the two points at the 'minimum stress. Curve III followed immediately, a well-defined and definite hysteresis loop. While at each load one point only is plotted, it should be mentioned that, if given time, noticeable creep will occur at the higher stress points on the increasing stress branch. On the descending branch, however, no observable creep occurs. After curve III was completed, the crystal was loaded to 50 g/mm² and various observations of creep made. It was then left under the minimum load for 22 hours, following which curve IV resulted. The first, straight part of the curve is, within the accuracy of observation, identical in slope to the "elastic" curve (I). Definite creep occurs, however, at the highest stress shown (2 minutes interval between the two readings), and the following run, curve V, again shows hysteresis. This hysteresis is not as

⁹ The first point on curve II and on the other curves is plotted at the correct stress value but the strain is set arbitrarily in order to separate or compare certain curves. Thus II and III are set together with the same initial strain whereas in fact III started from the last point observed in II. Curves IV and V are similarly arranged.

much as it is in curve III and the average slope is nearer that of the elastic curve.

The results depicted in Fig. 6 are very similar to those just described. Curve VI is an elastic curve taken after the crystal had had quite a few runs taken on it but had then rested for 10 days. Points were taken on this curve for both increasing and decreasing loads below 60 g/mm², and no observable hysteresis was found. At the highest stress shown the results described in connection with curve I were duplicated. The permanent set after unloading was $\Delta l/l = 1.7 \times 10^{-4}$. The lower part (hysteresis loop) of curve VII was taken next, and then the complete curve with repetition of the upper branch. In this curve the time intervals between successive points at the same stress are 1 minute, except for 2 minutes at 40.2 g/mm². After taking the upper part of curve VII and returning to the minimum load, the lower part (hysteresis loop) was duplicated. Curves VIII and IX were taken one and three days, respectively, after the data for curve VII, with no other runs intervening. The gradual recovery of the crystal is obvious. In both these curves the ascending branch is almost straight and parallel to the "elastic" curve (VI). Curve X shows three successive runs on a hysteresis loop like that of curve VII with points indicated by circles (O), crosses (+), and triangles (Δ), and taken in that order. These sets of observations agree within the experimental accuracy and a definite cyclic state seems to be indicated.

Creep data of the same type as that shown by the succession of points at each stress in curve VII, are represented in a different way in Fig. 7. These data were taken when the crystal was in a strained condition which was produced by starting with the crystal in an elastic state and then going beyond the elastic limit to a permanent set of $\Delta l/l = 0.8 \times 10^{-4}$. For each curve, except XI, the strain at the point marked A is arbitrarily plotted as zero although it is really the strain reached at the end of the preceding run (curve of next lower number). At A the increment to the load commenced and was completed several seconds before taking the reading at B, which is 20 seconds after A. The pronounced "toe" on curve XV shows the initial slowness

From the foregoing one may conclude that a zinc crystal of this orientation,¹⁰ when sufficiently rested (self-annealed), has an elastic limit which is very well defined by the simultaneous occurrence of the end of the Hooke's law region and the point at which marked and rapid creep occurs. This limit is of the type previously reported for lead crystals⁶ but the subsequent behavior of the zinc is very different from that of lead when the limit is passed and creep occurs. The zinc is now in an inelastic state in which Hooke's law fails both for increasing and decreasing stress and in which marked hysteresis occurs. It must be emphasized that the hysteresis loop cannot be wholly accounted for by assuming creep while the load on the specimen is being increased since one should then find, at the end of the cycle, a permanent set equal to the accumulated deviation from the elastic curve. This inelastic state gradually disappears with the self-annealing of the zinc. Before complete recovery the material may return to almost the original elastic state, with no or very little hysteresis, but with an elastic limit markedly lower than in the completely recovered state. Something like seven or eight days rest at room temperature seem to be needed for the limit to rise to the value of curves I and VI.

During the whole set of experiments on crystal No. 6 the total accumulated permanent set was only $\Delta l/l = 11 \times 10^{-4}$, or about 0.1 percent. The specimen showed no strain nor slip bands on its surface. A careful measurement of Young's modulus after an appropriate rest interval yielded $E = 11.6 \times 10^{11}$ dynes per cm^2 . This is about 3 percent lower than the initial value of 12.0×10^{11} dynes/ cm^2 , corresponding to the crystal in the state of curve I, Fig. 5.¹¹

¹⁰ More recent work at the University of Iowa has confirmed the behavior described above for crystals of orientations above 45° but has shown that crystals of orientation lower than 45° act somewhat differently.

¹¹ The critical reader will note that the slope of this curve does not correspond to the stated value of Young's modulus. This is because the curves were all plotted with the stress computed from a provisional value of cross section. The cross section was only obtained accurately at the conclusion of the work by cutting out and weighing the actual 9-cm length on which observations were made. It did not seem worth while to correct all the curves since the error in stress is small and is the same for all the curves.

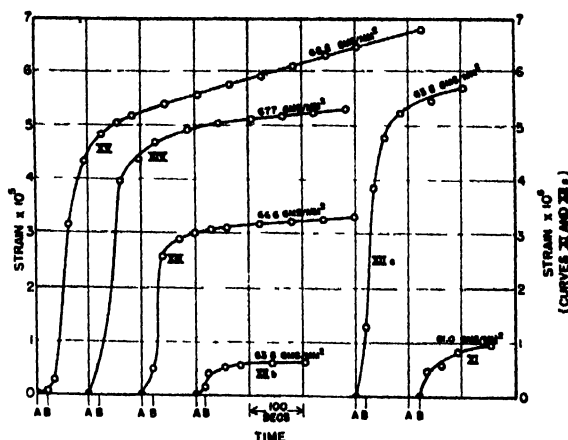


FIG. 7. Crystal No. 6: creep under successively greater loads. At *A* the load is that shown on the curve of next lower number. Between *A* and *B* the load is raised to the value shown on the curve and thereafter kept constant. The loading is complete several seconds before *B*. The interval *AB* is twenty seconds. Time increases from left to right with the lowest point on each curve at *t*=0. The initial strain for each curve is arbitrarily set equal to zero.

DISCUSSION

Some of the hitherto puzzling phenomena observed in connection with internal friction of zinc crystals may receive provisional and qualitative explanations in terms of the slow speed observations. If a zinc crystal may be put into the inelastic state by accidental strains incurred during handling, mounting for measurement, etc., then such a crystal would be expected to have a high decrement (as measured at the minimum stress amplitude). With some rest this decrement should decrease, but it might still rise with increasing stress amplitude until rested sufficiently for a complete return to the elastic state. When this state is reached the decrement at low stress amplitudes should be a minimum and the decrement curve should stay substantially level until at least the static elastic limit is reached. Indeed this point might be exceeded somewhat since it is the maximum stress amplitude which is plotted and this exists in only a very small portion of the crystal with less stress elsewhere. It is, however, rather obvious that in these experiments the decrement does not show any marked rise until the elastic limit has been greatly exceeded. Thus for crystal No. 4 (after the anneal, Fig. 2) a noticeable rise starts at about 230 g/mm² which is twice the elastic

limit to be expected¹² from the observations on crystal No. 6, of the same purity. For crystal No. 2 (Fig. 1) the decrement has only doubled when the stress is four times the elastic limit. It has not, however, been definitely shown that an inelastic state can be set up in a crystal of such low orientation¹⁰ by the means employed for the higher orientation crystals. Crystals Nos. 3 and 5 show level decrements to very much higher stresses. On account of their greater impurity the elastic limit may well be much higher than for crystal No. 6. No statement can therefore be made as to how much the elastic limit is exceeded in these crystals.

The observation that the decrement may remain fairly constant well beyond the static elastic limit seems to be connected with the initial slow rate of transient creep previously mentioned. With a rapidly alternating stress it seems likely that sufficient plastic deformation to cause an appreciable hysteresis loop, and a corresponding increase in the decrement, does not have time to occur. If the presence of dislocations is essential for creep and if the rate of creep depends on the number present,¹³ then both the rapid and slow speed experiments seem to point to the necessity of supposing that some sort of incubation period exists for generation of the many more dislocations needed to give either a rapid rise in decrement or the greatly increased rate of creep shown by the steep parts of the curves of Fig. 7.

If, however, the rate of creep depends on the rate of generation of dislocations it must be assumed that these are generated only at a low rate for the first small time interval following an increase in stress.

The difference found here between the two pairs of crystals of different purity may explain the behavior of the very pure crystals of reference 1. Such pure material may have so low an elastic³ limit that it is just about reached when

the stress amplitude is the lowest at which the measurements can be made. On this hypothesis the crystals would be in the inelastic state when mounting in the apparatus was completed, since strains due to handling would be unavoidable. The decrements observed immediately would be high, but would decline with time as the crystals annealed and approached the elastic state. The decrements could not be expected to stay level over any considerable range of stress amplitude, however.

Miller's¹⁴ experiments bear out some such supposition as to the very low elastic limit of very pure zinc since he obtained creep with a resolved shear stress equal to about one-third of that at the elastic limit of crystal No. 6. His creep rate (averaged for the first 80 hours) is roughly one-tenth of the steady creep rate taken from the last part of curve XV in Fig. 7. Such a rate would be well within observation with the present apparatus and definitely does not occur for a well-annealed zinc crystal of 99.99 percent purity. Of course if Miller's crystal was in the inelastic state, due to handling, etc., both the rate of creep and the stress at which it occurred seem possible without assuming a vanishingly low elastic limit. The change to a much faster rate of creep, however, after the strain had reached 5×10^{-4} seems to point to the onset of an inelastic state at this time rather than to its being present earlier.

Attempts to compare creep rates and internal friction have been made by Seitz and Read.¹⁵ If the mean transient creep rate taken from the steep parts of the curves of Fig. 7 is used, with a corresponding stress of 65 g/mm², a decrement of about 10^{-8} results. While this is of the right order of magnitude, the result must be regarded with some suspicion if there is any validity in the preceding speculations, in which it has been assumed that in the region of level decrement this rapid transient creep does not occur.

The writers are indebted to Mr. George Kuck for making the spectrographic analyses.

¹² Computed on the basis of the same resolved shear stress.

¹³ F. Seitz and T. A. Read, *J. App. Phys.* 12, 470 (1941). See in particular pp. 471 and 478.

¹⁴ See Fig. 40, reference 12.

¹⁵ Reference 12, Eqs. (4) and (5).

Design of Broad Band I. F. Amplifier. Part II

RICHARD F. BAUM

Raytheon Manufacturing Company, Waltham, Massachusetts

(Received May 13, 1946)

This paper is a continuation of a previous article on the design of broad band I.F. amplifiers of the stagger-tuned type. The analysis is extended to unrestricted band width. It is shown that an exact solution is possible for either an oscillatory or a monotonic response, with this result: 1. The necessary number of stages (l) for a given minimum attenuation in the cut-off region depends again only on the gain tolerance and the desired response. 2. Attenuation minima and maxima again appear at frequencies within the band, easily located and dependent only on the number of stages. 3. The amplifier consists of a number of pairs of

circuits, which have the same figure of merit Q_n and tune at resonant frequencies f_{0n} disposed in geometric symmetry around the middle band frequency f_0 thus making f_{0n} equal to $\delta_n f_0$ and f_0/δ_n , respectively. Formulas for Q_n and δ_n are derived. The circuit impedances are calculated from a prescribed gain or from the maximum obtainable gain. A formula for the maximum gain band width product is given. A comparison is made between the performance of amplifiers with monotonic and oscillatory response. A practical example is worked out.

(A) INTRODUCTION

IN a previous paper on the design of broad band I. F. amplifiers,¹ a method of calculating the circuit parameters was outlined, which leads to simple formulas both in the case of monotonic or oscillatory amplifier response. The formulas derived there give a very good insight into the influence and mutual dependence of the given and the calculated design quantities. This simplicity was achieved at the expense of band-width. It was necessary to introduce an approximation for the expression of the detuning ϵ . By replacing its exact value

$$\epsilon = f/f_0 - f_0/f$$

by its approximate value

$$\epsilon = 2(f - f_0)/f_0,$$

this approximation restricted the validity of all derived formulas to a band width not larger than say one-quarter the center frequency f_0 .

It is obvious that a design procedure permitting the exact calculation without restriction of band width would be welcome. This problem was successfully solved by Wallman² for the case of a monotonic response curve. In this paper the calculation is extended to the oscillatory case, which is often preferred to the monotonic case for its higher gain band width product and steeper cut-off slope.

Any particular desired response curve is substantially described by 4 points. Two of them define a band width within which the gain does not vary more than a definite prescribed amount. This amount is small and makes the gain in this region essentially constant. The two other points prescribe a certain minimum attenuation which has to be reached at certain frequencies outside of the pass band. The amount of required attenuation at these points as well as the admitted gain variation are given quantities for any particular design problem. In some cases two points may coincide with the half-power points; but formulas which are not restricted to the use of the half-power points are obviously more general and convenient.

Wallman has shown that by arranging the circuits in pairs of same Q_n and with resonance frequencies f_{01} and f_{02} symmetrically disposed around the middle frequency f_0 so as to make $f_{01} = \delta_1 f_0$ and $f_{02} = f_0/\delta_1$ a monotonic response curve of the form $1 + (Q\epsilon)^{2l}$ can be obtained, from which the individual Q_n and δ_n can be derived by factorization. In the present paper it shall be shown that the most general expression is a polynomial in ϵ of highest power $k_m = 2l$, containing only even powers of ϵ . The same line of reasoning is applied as in the previous paper; the polynomial is identified with an expression derived from Tschebyscheff's polynomial, the roots of which have been calculated. They furnish the Q_n and δ_n for the oscillatory case.

The same designations are used, as in the first

¹ R. F. Baum, J. App. Phys. 17, 519 (1946).

² H. Wallman, "Stagger tuned I. F. amplifiers," M.I.T. Report 524, February 1944.

paper, but some of the quantities will have to be defined in accordance with the more rigorous requirements of the present calculation. Because of the geometric symmetry, all corresponding frequencies will have the middle band frequency as their geometric mean, for instance the upper and lower band limits, the two points of prescribed attenuation outside the band, pairs of maximum or minimum gain, etc. The geometric symmetry of the response curve suggests further the introduction of a logarithmic frequency scale.

As the line of reasoning is very similar to that of the first paper, the deduction of formulas and the arrangement of paragraphs proceeds in the same sequence. In order to distinguish between the formulas of the first and second part the formulas of the second part have an additional index 2.

The introduction of a reference circuit, as in the first part, proves of only restricted value in the present case as no great simplification of formulas can be obtained.

(B) POSSIBLE ATTENUATION CURVES

With the same designations as in the first part, the gain of a single-tuned stage is:

$$G_n = g_n R_n / (1 + jQ_n \epsilon_n) \quad (2-1)$$

and its absolute value

$$|G_n| = g_n R_n / (1 + Q_n^2 \epsilon_n^2)^{1/2} \quad (2-2)$$

The exact expression for the detuning ϵ_n is

$$\epsilon_n = f/f_{0n} - f_{0n}/f \quad (2-3)$$

Again a reference circuit is introduced, which resonates at mid-band frequency f_0 and for which*

$$\epsilon = f/f_0 - f_0/f \quad (2-4)$$

A logarithmic coordinate system is to be used. The variables become

$$\beta_n = \lg(f/f_{0n}), \quad \beta = \lg(f/f_0), \quad (2-5)$$

thus

$$\epsilon_n = 2 \sinh \beta_n, \quad \epsilon = 2 \sinh \beta. \quad (2-6)$$

The resonance frequency f_{0n} of the n th circuit

* Quantities without index n refer to the reference circuit.

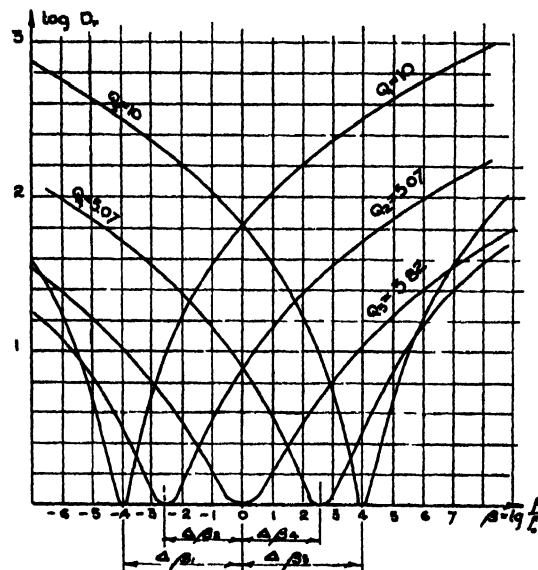


FIG. 1. Response curves of 5 stages. The center stage has the lowest Q and tunes at the geometric mean frequency f_0 . The other stages are arranged in pairs of equal Q_n and with resonance points shifted the same distance above and below f_0 on a logarithmic frequency scale.

shall be determined by a factor δ_n :

$$f_{0n} = \delta_n f_0. \quad (2-7)$$

Its logarithm is

$$\Delta\beta_n = \lg \delta_n, \quad (2-8)$$

which makes

$$\epsilon_n = 2 \sinh(\beta - \Delta\beta_n). \quad (2-9)$$

In the new coordinate system the origin corresponds to $f/f_0 = 1$, that is, to the mid-band frequency. The fact that the n th circuit resonates at $\delta_n f_0$ thus appears as a simple shift of its response curve by an amount of $\Delta\beta_n$ to the right (Fig. 1). The gain becomes

$$|G_n| = g_n R_n [1 + Q_n^2 4 \sinh^2(\beta - \Delta\beta_n)]^{-1/2}, \quad (2-10)$$

which shows that the response curves are symmetric with respect to the point $\beta = \Delta\beta_n$.

The index $n=1$ again is given to the circuit resonating at the low frequency end of the band and the index $n=l$ to the circuit resonating at the upper end. The indices $n=2$ and $n=l-1$ are reserved for the two adjacent stages, etc.

The gain of l stages is

$$|G_n| = \prod_{n=1}^{n=l} g_n R_n [1 + 4Q_n^2 \sinh^2(\beta - \Delta\beta_n)]^{-1/2}. \quad (2-11)$$

At the mean frequency, β equals zero and the

gain becomes:

$$|G_0| = \prod_{n=1}^{n=t} g_n R_n [1 + 4Q_n^2 \sinh^2 \Delta\beta_n]^{-1}. \quad (2-12)$$

If all stages should tune to the same frequency the gain would be:

$$|G_0| = \prod_{n=1}^{n=t} g_n R_n. \quad (2-13)$$

The loss in gain due to the displacement of the resonant frequencies is thus

$$L^2 = \prod_{n=1}^{n=t} [1 + 4Q_n^2 \sinh^2 \Delta\beta_n], \quad (2-14)$$

and the gain can again be written in the form

$$|G_0| = (1/L) \prod_{n=1}^{n=t} g_n R_n. \quad (2-15)$$

The part of (2-10) which is a function of frequency defines all possible response curves and is:

$$D_P = \epsilon^{2d} = \prod_{n=1}^{n=t} [1 + 4Q_n^2 \sinh^2 (\beta - \Delta\beta_n)], \quad (2-16)$$

where d is the voltage attenuation in Nepers.

Now the circuits may be arranged in pairs of the same Q_n and of resonance frequencies disposed in geometric symmetry to both sides of the mid-band frequency.* On a logarithmic scale this corresponds to a shift of $\Delta\beta_n$ in the positive sense for one circuit and in the negative sense for the other. The response of such a pair is:

$$\begin{aligned} & [1 + 2jQ_n \sinh (\beta - \Delta\beta_n)] \\ & \quad \times [1 + 2jQ_n \sinh (\beta + \Delta\beta_n)] \\ & = 1 - 2Q_n^2 [\cosh (2\beta) - \cosh (2\Delta\beta_n)] \\ & \quad + j4Q_n \sinh \beta \cosh \Delta\beta_n \\ & = [1 + 4Q_n^2 \sinh^2 \Delta\beta_n] - 4Q_n^2 \sinh^2 \beta \\ & \quad + j4Q_n \sinh \beta \cosh \Delta\beta_n \\ & = A - 4Q_n^2 \sinh^2 \beta + j2BQ_n \sinh \beta. \end{aligned}$$

Reintroducing ϵ from (2-6) the response of one pair becomes

$$D_P^1 = [A - Q_n^2 \epsilon^2]^2 + Q_n^2 \epsilon^2 B^2, \quad (2-17)$$

where A and B are constants.

* This idea and the corresponding proof is contained in Wallman's paper, reference 2.

This equation shows that the response curve of a pair of circuits is a polynomial of the fourth power of the variable ϵ , and containing only even powers of ϵ . The response of all stages is a polynomial obtained by multiplication of several such polynomials and can therefore possess only even powers of ϵ . Its highest power k_m will be $k_m = 2t$ or twice the number of circuits. This fact is obviously not altered by the addition of one single circuit resonating at mid-band frequency f_0 , the response of which is quadratic in ϵ :

$$D_P = 1 + Q^2 \epsilon^2.$$

(C) THE REQUIRED RESPONSE CURVE

The result obtained checks closely with the one obtained in the first part of this paper, the only difference being that the variable ϵ now is defined by its exact value (2-4) instead of its approximate value according to Eq. (1-6). Thus the same reasoning may be applied to the present case as to the former:

It was shown that the required response curve could be represented in either of two forms:

$$D_r = |d_0| T_{2t}(\alpha) + (1 + d_0) \quad (2-18)$$

for the oscillatory case, or

$$D_r^* = 1 + 2d_0 \alpha^{2t} \quad (2-19)$$

for the monotonic case. In these expressions d_0 is the admissible gain variation within the pass band, and the variable α was defined by Eq. (1-20) with

$$\alpha = s/s_L = \epsilon/\epsilon_L. \quad (2-20)$$

This ϵ_L is that value which the variable ϵ takes at the ends of the band. This definition still applies to the present case although ϵ_L now is given by

$$\epsilon_L = f_L/f_0 - f_0/f_L, \quad (2-21)$$

where f_L is the upper frequency limit of the band. Setting

$$f_L = \delta_L f_0, \quad (2-22)$$

ϵ_L becomes

$$\epsilon_L = \delta_L - 1/\delta_L = 2 \sinh \beta_L, \quad (2-23)$$

where $\beta_L = \lg \delta_L$ and is the equivalent of ϵ_L on the logarithmic frequency scale.

The variable α thus becomes

$$\alpha = \epsilon/\epsilon_L = 2 \sinh \beta / \epsilon_L = \sinh \beta / \sinh \beta_L. \quad (2-24)$$

It was shown that the points of minimum attenuation within the band occur at points uniquely determined by the number l of stages employed and given (see Eq. (1-23)) by

$$\left. \begin{aligned} \alpha_n^{\min} &= \cos [(2m+1)\pi/2l] \\ m &= 0, 1, 2, \dots (l-1) \\ &= n-1 \end{aligned} \right\}. \quad (2-25)$$

To this corresponds a certain detuning $\epsilon_n^{\min}$, a certain frequency $f_n^{\min}$, and a certain shift $\beta_n^{\min}$ on the logarithmic scale, given by

$$\alpha_n^{\min} = \epsilon_n^{\min} / \epsilon_L = 2 \sinh \beta_n^{\min} / \epsilon_L \quad (2-26)$$

and

$$\beta_n^{\min} = \lg (f_n^{\min} / f_0) = \lg \delta_n^{\min}. \quad (2-27)$$

Furthermore, it was shown that it is possible to calculate the complex roots α_{0n} of (2-18) and (2-19). Their real and imaginary parts were designated with α_{Rn} and α_{In} , respectively:

$$\alpha_{0n} = \alpha_{Rn} + j\alpha_{In}. \quad (2-28)$$

Their values were (Eq. (1-34) and (1-35)) for the oscillatory case [with added index n]:

$$\alpha_{Rn} = \cosh (c/2l) \cos [(2m+1)\pi/2l], \quad (2-29)$$

$$\alpha_{In} = \sinh (c/2l) \sin [(2m+1)\pi/2l], \quad (2-30)$$

where (Eq. (1-32)):

$$c = \cosh^{-1} \frac{1+d_0}{d_0},$$

and for the monotonic case (Eqs. (1-38) and (1-39))

$$\alpha_{Rn}^* = [2d_0]^{-1/2l} \cos [(2m+1)\pi/2l], \quad (2-31)$$

$$\alpha_{In}^* = [2d_0]^{-1/2l} \sin [(2m+1)\pi/2l]. \quad (2-32)$$

The roots corresponding to α_{0n} are, in the logarithmic coordinate system, given by β_{0n} :

$$\alpha_{0n} = \epsilon_{0n} / \epsilon_L = 2 \sinh \beta_{0n} / \epsilon_L. \quad (2-33)$$

Now as α_{0n} is complex, β_{0n} obviously is complex also:

$$\beta_{0n} = \beta_{Rn} + j\beta_{In}. \quad (2-34)$$

This introduced into (2-33) furnishes, after separation of real and imaginary components:

$$\alpha_{Rn} = \frac{2}{\epsilon_L} \sinh \beta_{Rn} \cos \beta_{In}, \quad (2-35)$$

$$\alpha_{In} = \frac{2}{\epsilon_L} \cosh \beta_{Rn} \sin \beta_{In}. \quad (2-36)$$

These two equations can easily be solved for β_{Rn} and β_{In} . The solution is

$$\cosh (2\beta_{Rn}) = \frac{\sigma_n^2}{4} + \left[\left(\frac{\sigma_n^2}{4} \right)^2 + \frac{\tau_n^2}{2} + 1 \right]^{\frac{1}{2}}, \quad (2-37)$$

$$\cos (2\beta_{In}) = -\frac{\sigma_n^2}{4} + \left[\left(\frac{\sigma_n^2}{4} \right)^2 + \frac{\tau_n^2}{2} + 1 \right]^{\frac{1}{2}}, \quad (2-38)$$

where

$$\sigma_n^2 = \epsilon_L^2 (\alpha_{Rn}^2 + \alpha_{In}^2), \quad (2-39)$$

$$\tau_n^2 = \epsilon_L^2 (\alpha_{Rn}^2 - \alpha_{In}^2). \quad (2-40)$$

By subtraction and by multiplication of (2-37) and (2-38) it is found that between β_{Rn} and β_{In} the following relations must exist:

$$\frac{\sigma_n^2}{4} = \sinh^2 \beta_{Rn} + \sin^2 \beta_{In}, \quad (2-41)$$

and

$$\frac{\tau_n^2}{2} + 1 = \cosh (2\beta_{Rn}) \cos (2\beta_{In}). \quad (2-42)$$

The quantity σ_n introduced here for convenience obtains significance in the calculation of the gain. β_{Rn} is to furnish the resonance shifts $\Delta\beta_n$, and β_{In} is to furnish the Q_n of the individual circuits.

Equations (2-35) to (2-42) apply as well to the monotonic as to the oscillatory case, depending upon which roots, (2-31), (2-32), or (2-29), (2-30), are introduced.

(D) CALCULATION OF CIRCUIT PARAMETERS

a. Selectivities and Resonant Frequencies

By equalizing the roots given by Eq. (2-16)

$$D_F = \prod_{n=1}^{n=l} [1 + 4Q_n^2 \sinh^2 (\beta_{0n} - \Delta\beta_n)] = 0, \quad (2-43)$$

with the roots β_{0n} as calculated in the previous paragraph, all circuit parameters may be calculated. By setting each factor of (2-43) equal to zero, its root $\beta_{0n} = \beta_{Rn} + j\beta_{In}$ is found from:

$$\begin{aligned} 1 + 4Q_n^2 \sinh^2 (\beta_{0n} - \Delta\beta_n) &= 0, \\ \sinh (\beta_{Rn} + j\beta_{In} - \Delta\beta_n) &= j/2Q_n. \end{aligned} \quad (2-44)$$

The right side of this equation is imaginary; thus on the left side the real component has to equal zero, or

$$\Delta\beta_n = \beta_{Rn}. \quad (2-45)$$

This leaves

$$\sin \beta_{In} = 1/2Q_n. \quad (2-46)$$

Equations (2-45) and (2-46) determine the simple relations between Q_n and $\Delta\beta_n$ (the selectivities and resonant frequencies) and the quantities βR_n and βI_n , as given by Eq. (2-37) and (2-38). Thus the problem is solved in principle. Whereas in the approximate case (treated in Part I) the Q_n and the resonant frequencies depended separately on the real or imaginary root components αR_n and αI_n , respectively, it is seen, that in the rigorous calculation they depend on both components in a rather involved manner. For this reason the result cannot be simply stated. Thus the importance of the approximate solution lies partly in the way it indicates in simple formulas the trend of the phenomena and the interconnection of the design parameters, whereas the final design can be based on the exact solution.

For both the oscillatory and the monotonic response, expressions (2-39) and (2-40) can be simplified:

In the oscillatory case with (2-29) and (2-30):

$$\sigma_n^2 = \frac{\epsilon_L^2}{2} \{ \cosh(c/t) + \cos[(2m+1)\pi/t] \},$$

$$= \epsilon_L^2 \{ \sinh^2(c/2t) + \cos^2[(2m+1)\pi/2t] \}, \quad (2-47)$$

$$\tau_n^2 = \frac{\epsilon_L^2}{2} \{ 1 + \cosh(c/t) \cos[(2m+1)\pi/t] \}, \quad (2-48)$$

in the monotonic case with (2-31) and (2-32):

$$\sigma_n^{*2} = \epsilon_L^2 [2d_0]^{-1/t}, \quad (2-49)$$

$$\tau_n^{*2} = \epsilon_L^2 [2d_0]^{-1/t} \cos[(2m+1)\pi/t]; \quad (2-50)$$

$$m = n-1 = 0, 1, 2, \dots, (t-1).$$

In the beginning the gain was written in terms of a reference circuit, which was supposed to resonate at mid-band frequency f_0 and has a Q_n equal Q (without index). This circuit requires the following solutions (from (2-45) and (2-46)):

$$\beta R_n = \Delta\beta_n = 0$$

and

$$\sin \beta I_n = 1/2Q$$

which are obtained by setting

$$(2m+1)\pi/2t = \pi/2.$$

This gives, in the oscillatory case (Eqs. (2-29),

(2-30), (2-35), and (2-26)):

$$\alpha_R = 0, \quad \alpha_T = \sinh(c/2t) = 1/Q\epsilon_L, \quad (2-51)$$

and in the monotonic case (Eqs. (2-31), (2-32), (2-36), and (2-37)):

$$\alpha_R^* = 0, \quad \alpha_T^* = [2d_0]^{-1/2t} = 1/Q\epsilon_L. \quad (2-52)$$

These relations correspond exactly to formulas (1-45) and (1-49) of the first part. They obviously give a very simple relationship between the band width ϵ_L , the admissible gain variation α_0 , and the Q of the reference circuit, which in the case of an uneven number of stages coincides with the middle stage. The corresponding value of σ becomes in both cases

$$\sigma = \sigma^* = 1/Q. \quad (2-53)$$

From (2-49) furthermore (in the monotonic case):

$$\sigma_n^* = \sigma^* = 1/Q \text{ constant for all stages.}$$

b. The Minimum Number of Circuits

Nothing has to be added to the corresponding chapter of Part I. All formulas derived there remain valid in the present case and this without restriction for the values of the quantity α_m appearing in formulas (1-57) to (1-64), the value of which is now defined by two frequencies, f_{m1} and f_{m2} ($f_{m1} \cdot f_{m2} = f_0^2$) at which the attenuation d should attain at least a prescribed minimum value d_m . Writing

$$\epsilon_m = f_{m1}/f_0 - f_0/f_{m1}, \quad (2-54)$$

and

$$\epsilon_L = f_L/f_0 - f_0/f_L$$

as in (2-20), α_m is given by

$$\alpha_m = \epsilon_m/\epsilon_L. \quad (2-55)$$

c. Impedances and Gain

The loss in gain due to the staggering of the resonance frequencies was (Eq. (2-14)):

$$P = \prod_{n=1}^{n=t} [1 + 4Q_n^2 \sinh^2 \Delta\beta_n]$$

$$= \prod_{n=1}^{n=t} 4Q_n^2 \left[\left(\frac{1}{2Q_n} \right)^2 + \sinh^2 \Delta\beta_n \right].$$

By introduction of Eqs. (2-45), (2-46), and (2-41)

this expression reduces to

$$l = \prod_{n=1}^{n-1} Q_n \sigma_n. \quad (2-56)$$

The product $\sigma_n Q_n$ of each circuit determines its loss in gain (or attenuation) at mid-band frequency. This loss, of course, is the same for each circuit of a circuit pair for reasons of symmetry, but is different for each pair.

The product given by Eq. (2-56) thus consists of a series of pairs of equal factors (squares). A single stage resonating at mid-band frequency obviously does not contribute to the gain loss, because, according to Eq. (2-53): $\sigma = 1/Q$.

In the case of monotonic response the product (2-56) is simplified because σ_n is the same (see Eq. (2-49)) for all circuit pairs:

$$l^* = \epsilon_L^i [2d_0]^{-1} \prod_{n=1}^{n-1} Q_n. \quad (2-57)$$

The expressions for the gain loss due to resonance staggering can be written also in terms of the Q of the reference circuit by introducing σ_n and Q from (2-51) or (2-52). They become

$$l = \prod_{n=1}^{n-1} \frac{Q_n}{Q} \left[1 + \frac{\cos^2 [(2m+1)\pi/2i]}{\sinh^2 (c/2i)} \right]^{\frac{1}{2}} \quad (2-58)$$

for the oscillatory case, and

$$l^* = \prod_{n=1}^{n-1} \frac{Q_n}{Q} \quad (2-59)$$

for the monotonic case.

The last equation shows that in the monotonic case the gain loss l^* increases for each stage in proportion to the increase of Q_n over the Q of the reference circuit. But according to Eq. (2-58) in the oscillatory case this loss increases more rapidly and differs from circuit pair to circuit pair.

The loss in gain due to the staggering of resonance frequencies is, of course, totally or partly compensated by the increase in stage gain due to the increase in Q_n as will be apparent from the following gain formulas.

The total gain according to Eq. (2-15) was

$$|G_0| = (1/l) \prod_{n=1}^{n-1} g_n R_n,$$

and with (2-56)

$$|G_0| = \prod_{n=1}^{n-1} g_n R_n / Q_n \sigma_n \quad (2-60)$$

$$= \prod_{n=1}^{n-1} g_n / C_n \omega_{0n} \sigma_n. \quad (2-61)$$

As any circuit pair resonates at frequencies the product of which equals the square of the mid-band frequency, one may replace, in the product, ω_{0n} by ω_0 . Furthermore, considering only the case of stages with the same capacities $C_n = C$, and the same transconductance $g_n = g$, the gain becomes:

$$|G_0| = \left[\frac{g}{C\omega_0} \right]^i \prod_{n=1}^{n-1} \frac{1}{\sigma_n}. \quad (2-62)$$

Writing this expression for just one pair of circuits it is easily seen that in the monotonic case the average gain per circuit pair equals the gain of the reference (or middle) circuit.

The average gain per stage is

$$|\bar{G}_0| = \frac{g}{C\omega_0} \left[\prod_{n=0}^{n-1} \frac{1}{\sigma_n} \right]^{1/i}. \quad (2-63)$$

The total band width (with Eqs. (2-22) and (2-23)) is

$$BW = \delta_i f_0 - \frac{f_0}{\delta_L} = f_0 \epsilon_i, \quad (2-64)$$

thus the average gain band with product becomes

$$|\bar{G}_0| BW = \left(\frac{g}{\pi C} \right) \cdot \frac{\epsilon_L}{2} \left[\prod_{n=1}^{n-1} \frac{1}{\sigma_n} \right]^{1/i}. \quad (2-65)$$

For the monotonic case the root equals Q and with (2-52) the expression becomes

$$|\bar{G}_0^*| BW = \left(\frac{g}{\pi C} \right) \cdot \frac{1}{2} [2d_0]^{1/2i} \quad (2-66)$$

which is exactly the same as the expression derived for the approximate calculation (see Eq. (1-78)).

The factor in parenthesis again is recognized as the theoretical maximum gain band width product and again an efficiency factor may be defined by:

$$\mu = \frac{\epsilon_L}{2} \left[\prod_{n=1}^{n-1} \frac{1}{\sigma_n} \right]^{1/i}, \quad (2-67)$$

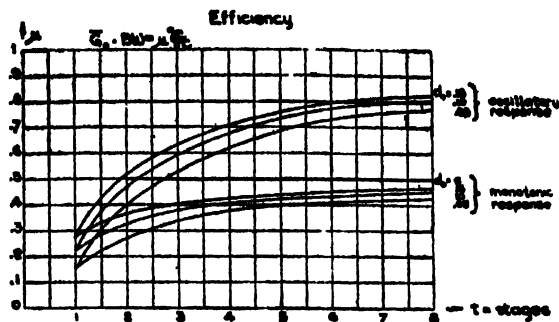


FIG. 2. Efficiency of amplifiers with oscillatory and monotonic response.

or with (2-47):

$$\mu = \frac{1}{2} \prod_{n=1}^{n=t} \left\{ \sinh^2(c/2t) + \cos^2[(2m+1)\pi/2t] \right\}^{-1/2} \quad (2-68)$$

and in the monotonic case:

$$\mu^* = \frac{1}{2} [2d_0]^{1/2t}, \quad (2-69)$$

the latter approaching rapidly $\frac{1}{2}$ with increasing the number of stages.

The importance of the quantity σ_n in all gain calculations is apparent. With reference to Eq. (2-62) one can say that σ_n defines the relative contribution of each individual stage to the total gain. In order to facilitate the gain calculation, the value of μ and μ^* is plotted in Fig. 2 as a function of t for $d_0 = .05$, $d_0 = .1$, and $d_0 = .15$.

(E) COMPARISON OF MONOTONIC AND OSCILLATORY RESPONSE

The deduced formulas permit the comparison of the performance of amplifiers of monotonic and oscillatory response.

For a given response characteristic, that is for a given maximum gain tolerance d_0 within the pass band and a given minimum attenuation d_m above a given frequency, a monotonic amplifier requires more stages than the oscillatory one. The ratio of their number is, with Eqs. (1-58) and (1-63):

$$\frac{t^*}{t} = \frac{h^*}{h} \frac{\cosh^{-1} \alpha_m}{\lg \alpha_m}. \quad (2-70)$$

The first factor almost equals one, so that approximately

$$\frac{t^*}{t} = \frac{\cosh^{-1} \alpha_m}{\lg \alpha_m}, \quad (2-71)$$

thus if the minimum attenuation is defined anywhere between $\alpha_m = \alpha_m^* = 1.1$ and 2, this ratio varies between 4.5 and 1.9.

For a given gain tolerance within the pass band, and with the same number of stages, the monotonic amplifier has a much flatter cut-off characteristic than the oscillatory one. Thus by setting $t/t^* = 1$ in Eq. (2-71) the frequencies at

TABLE I: Calculation of the Q_n and resonance frequencies.

stage 1 and 5 2 and 4 3 Eq.	m 0 and 4 1 and 3 2	$(2m+1)\pi/t = \theta$ 36° and 324° 108° and 252° 180°	$\cos \theta$.809 -.309 -1	$\cosh c/t + \cos \theta$ 2.006 .888
stage 1 and 5 2 and 4 3 Eq.	$\cosh c/2t \cdot \cos \theta$.97 -.37	σ_n^2 .7 .310	$\sigma_n^2/4$.175 .0776	$(\sigma_n^2/4)^2$.0305 .006
		(2-47)		$r_n^2/2$.343 .109
				(2-48)
stage 1 and 5 2 and 4 3 Eq.	σ .837 .557 .262 (2-53)	$\cosh 2\beta R_n$ 1.345 1.136	$\cos 2\beta T_n$.995 .980	$2\beta R_n$.797 .516
		(2-37)	(2-38)	$2\beta T_n$ 5.7° 11.3°
				$\Delta\beta_n$.398 .258 0 (2-45)
stage 1 and 5 2 and 4 3 Eq.	$\sin \beta T_n$.050 .0984	Q_n 10 5.07 3.82 (2-46) (2-53)	δ_n 1.49 1.29	f_m 17.9 and 8.05 mc 15.5 and 9.3 mc 12 mc (2-7)
			(2-8)	

which a given attenuation is reached are given by

$$\alpha_m^* = \alpha_m + (\alpha_m^2 - 1)^{1/2}. \quad (2-72)$$

Within a range of $1.1 < \alpha_m < 2$ the corresponding α_m^* lies between 1.55 and 3.7.

For a given minimum attenuation and a given number of stages the gain variation of the monotonic amplifier is obviously larger.

The average gain band width product, with a given number of stages and a given gain variation, is larger for the oscillatory amplifier by a factor: (Eqs. (2-68) and (2-69))

$$\frac{|\tilde{G}_0| \cdot BW}{|\tilde{G}_0^*| \cdot BW} = \frac{\mu}{\mu^*} \quad (2-73)$$

The value of μ and μ^* was plotted in Fig. 2.

Whereas the value of μ^* is somewhat smaller than $\frac{1}{2}$ the value of μ is smaller than 1. The limits of μ can be seen to be

$$\mu \rightarrow \frac{1}{2} \sinh(c/2t) \quad \text{for} \quad \sinh(c/2t) \gg \cos \frac{(2m+1)\pi}{2t}$$

and

$$\mu \rightarrow 2^{-1/t} \quad \text{for} \quad \sinh(c/2t) \rightarrow 0.$$

The first case corresponds to a very small gain tolerance and a small number of stages. It falls to values below one-half, but is still larger than the corresponding μ^* of the monotonic amplifier.

The other case corresponds to a large gain tolerance and a large number of stages. The

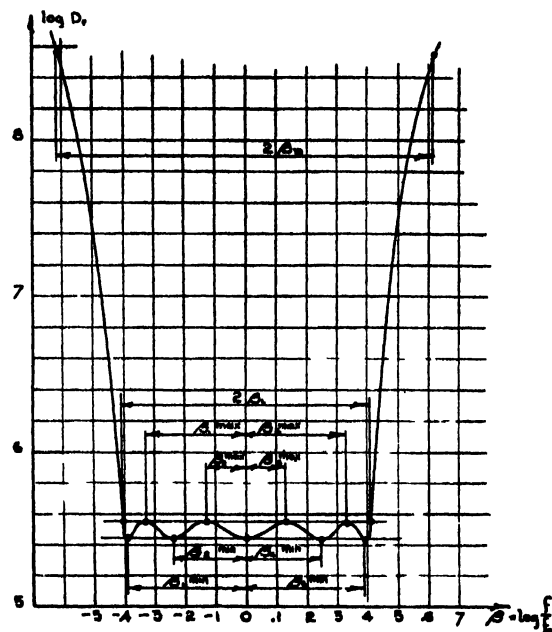


FIG. 3. The resultant response curve shows an oscillatory behavior within the pass band $2\beta_L$, defined by a given gain tolerance d_0 . The attenuation exceeds a prescribed minimum value d_m outside a given range $2\beta_m$.

average efficiency increases with an increasing number of stages; the total gain band width product tends towards $\frac{1}{2}$, whereas in the monotonic case the average gain band width product tends toward $\frac{1}{2}$.

From the standpoint of amplitude response, which was considered exclusively, the oscillatory amplifier is superior to the monotonic one in every respect.

TABLE II.

m	$\frac{(2m+1)\pi}{2t} = \frac{\theta}{2}$	$\cos \frac{\theta}{2}$	$\alpha_m^{\min}$	$\beta_m^{\min}$	$\delta_m^{\min}$	$f_m^{\min}$
0 and 4	18° and 162°	.951	.794	.387	1.473	8.16 and 17.7 mc
1 and 3	54° and 126°	.588	.490	.242	1.274	9.42 and 15.3 mc
2	90°	0	0	0	1	12 mc
formula		(2-25)	(2-26)	(2-26)	(2-27)	

47

TABLE III.

m	$\frac{m\pi}{t}$	$\cos \left(\frac{m\pi}{t} \right)$	$\alpha_m^{\max}$	$\beta_m^{\max}$	$\delta_m^{\max}$	$f_m^{\max}$
1 and 4	36° and 144°	.809	.675	.331	1.392	8.64 and 16.7 mc
2 and 3	72° and 108°	.309	.258	.129	1.138	10.55 and 13.7 mc

(F) EXAMPLE

A stagger tuned I.F. amplifier is to be designed having a band width from 8 to 18 mc, and having the maximum gain obtainable with 6AK5 tubes, with a gain variation of .1 Nepers within the pass band, with a power attenuation of at least 30 db at 22 mc and the corresponding lower frequency. An oscillatory response is desired. The given quantities are:

- $d_0 = .1$ Neper, the admissible gain variation.
- $\epsilon^{2d_0} = 1000$, the required power attenuation.
- $f_L = 18$ mc and 8 mc, respectively, the upper and lower band limits.
- $f_0 = (8.18)^{1/2} = 12$ mc, the geometric mean frequency,

$$f_n = 22 \quad \text{and} \quad \frac{12^2}{22} = 6.55 \text{ mc}$$

the frequencies at which the power drops to the required level. To f_L and f_m correspond detunings:

$$\epsilon_L = .83$$

and

$$\epsilon_m = 1.284$$

which correspond on the logarithmic scale to

$$\beta_L = .407$$

and

$$\beta_m = .607$$

(see Fig. 3). Thus

$$\alpha_m = \epsilon_m / \epsilon_L = 1.542 \quad \text{and} \quad \cosh^{-1} \alpha_m = 1.$$

The minimum number of stages (from (1-58) and (1-59)):

$$2t \geq h / \cosh^{-1} \alpha_m$$

$$\cosh h = [\epsilon^{2d_0} - (1 + d_0)] / d_0$$

$$\geq 10000$$

$$h = 9.8.$$

$$\geq 9.8$$

In order to obtain the prescribed response curve, at least 5 stages are required. Calculation of the hyperbolic expressions:

$$c = \cosh^{-1} \frac{1 + d_0}{|d_0|} = \cosh^{-1} 11 = 3.09,$$

$$c/2t = \frac{3.09}{10} = .309,$$

$$c/t = .618,$$

$$\cosh (c/2t) = 1.048,$$

$$\sinh (c/2t) = .314,$$

$$\cosh (c/t) = 1.197.$$

The resonance frequencies f_{0n} and the Q_n of the circuits have been calculated in Table I. The result has been plotted on a logarithmic scale in Fig. 1. By addition of the 5 individual response curves the total amplifier response of Fig. 3 is obtained which shows the required oscillatory behavior. The location of minimum attenuation is shown in Table II.

The points of maximum attenuation can be calculated in an analogous way from:

$$\alpha_m^{\max} = \cos \left(\frac{m\pi}{t} \right), \quad m = 1, 2, 3 \cdots (t-1), \quad (2-74)$$

which follows from inspection of Fig. 3 of Part I, (Table III).

The corresponding maximum and minimum values of attenuation within the band turn out to be 5.46 and 5.54. The former checks with the loss in gain if calculated from Eq. (2-58). The difference of .08 checks with the admissible gain variation, as $\log_{10} (1.1)^2 = .08$. (All curves are power response curves.) The minimum attenuation of 30 db = 3 is reached within the prescribed frequency limit $\beta_m = \pm .607$ with an attenuation of 8.57.

Assuming a transconductance of $g = 5000$ mhos and a minimum capacity of $12 \mu\mu\text{f}$ for each stage, the theoretical maximum gain band width product is:

$$9/\pi C = 133 \text{ mc.}$$

The efficiency factor is calculated from Eq. (2-66) with

$$\mu = \frac{.833}{2} / (.837^2 \times .557^2 \times .262)^{1/2} = .73.$$

The average gain band width product is thus:

$$|\bar{G}_0| BW = 133 \times .73 = 98.5 \text{ mc.}$$

With a band width of 10 mc the average gain

becomes

$$|\bar{G}_0| = 98.5/10 = 9.85,$$

and the total gain:

$$|G_0| = 9.85^4 = 92,400.$$

This compares very favorably with an amplifier of monotonic response whose best efficiency factor is .5. This gives a maximum gain for 5 stages (with 10 mc band width) of only 13,000.

CONCLUSION

It has been shown that the stagger tuned amplifier problem with oscillatory response can be exactly solved mathematically. The solution is obtained by the same means employed for the solution of the approximate case (with restricted band width), that is—by calculation of the roots of either a parabolic or a Tschebyscheff polynomial.

The Numerical Solution of Laplace's Equation in Composite Rectangular Areas

MAX M. FROCHT*

Illinois Institute of Technology, Chicago, Illinois

(Received January 29, 1946)

An analytical method is developed which yields key values of a harmonic function U in rectangular figures in terms of certain simple combinations of known boundary values, which we call Laplacian perimeters and denote by P . From the key values all the remaining interior values can be directly and rapidly calculated. The basic ideas consist of the introduction of four elementary figures of variable lengths: $3(2+\delta)$, $3(3+\delta)$, $4(2+\delta)$, and $4(3+\delta)$, where $0 \leq \delta \leq 1$, for which the key values are found algebraically, Eqs. (9.1) to (9.8) inclusive, and of combining analytically these elementary cases to obtain continuous solutions for rectangles $(3 \times n')$ and $(4 \times n')$ for any number n' greater than 2. The general expression for any key value $(u_1)_{m \times n'}$ in a rectangle $(m \times n')$ is

$$(u_1)_{m \times n'} = C_1 P_1 + C_2 P_2 + \dots + C_i P_i + \dots + C_k P_k,$$

in which $C_1, C_2, \dots, C_k$ are constant coefficients. Analytical relations are established between the coefficients for the last key value u_k in a rectangle $(m \times n)$, where n is an integer, and the coefficients for the next key value u_{k+1} in the rectangle $m(n+\delta)$ or $m(n+\delta+1)$. Specific values of the coefficients are presented in tabular form for rectangles $(3 \times n')$ and $(4 \times n')$ with the aid of which all key values in such rectangles can be rapidly calculated. The tables also permit the calculation of any one interior key value by itself. A numerical example and further approximate formulas are given, and applications to composite rectangular areas such as angles, channels, etc., are discussed.

1. INTRODUCTION

NUMERICAL or approximate solutions of Laplace's equation

$$\partial^2 U / \partial x^2 + \partial^2 U / \partial y^2 = 0 \quad (1.1)$$

are generally based on the fact that the value of the function U at an arbitrary point O of the region R to which the function U applies, Fig. 1, must satisfy approximately the general equation†

$$\left(\frac{1}{ac} + \frac{1}{bd} \right) u_o = \frac{u_a}{a(a+c)} + \frac{u_b}{b(b+d)} + \frac{u_c}{c(c+a)} + \frac{u_d}{d(d+b)} \quad (a) \quad (1.2)$$

or

$$u_o = k_a u_a + k_b u_b + k_c u_c + k_d u_d, \quad (b)$$

in which u_a, u_b, u_c , and u_d are the values of the harmonic function U at four neighboring points, A, B, C , and D , respectively, with coordinates as shown in Fig. 1. Equation (1.2) will be referred to as the four-point influence equation, or more briefly the four-point equation.

The procedure in numerical solutions generally consists of drawing a rectangular, equally spaced network distant h apart, and adjusting the values at each point of intersection of the network lines (network points) until they satisfy the four-point equation. When that is attained, an approximate or numerical solution is said to have been found. The distance h between the network

* Professor of Mechanics.

† For a derivation see references 1 and 3 in bibliography at end of paper.

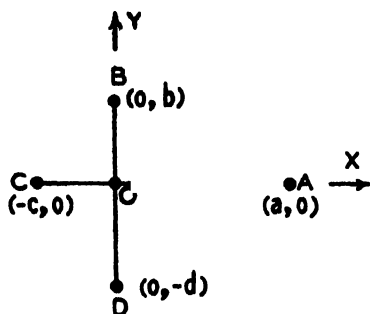


FIG. 1.

lines is called the lattice unit. The points A , B , C , and D may be located on the network lines, or on lines inclined 45° to them. In the first case we speak of normal neighbors, in the second case of diagonal neighbors. In the treatment which follows, both types are used.

Most of the interior network points will have equally spaced neighbors. For such points (1.2) reduces to

$$4u_o = u_a + u_b + u_c + u_d. \quad (1.3)$$

Equation (1.3) is generally called the Liebmann formula.

Another case of special interest is the case in which $a=b$, and $c=d$, Fig. 1. For this case (1.2) reduces to

$$u_o = \frac{c(u_a + u_b)}{2(a+c)} + \frac{a(u_c + u_d)}{2(a+c)}. \quad (1.4)$$

Assuming $a < c$ and $a/c = \delta$, this may be written as

$$u_o = \frac{(u_a + u_b)}{2(\delta+1)} + \frac{\delta(u_c + u_d)}{2(\delta+1)}, \quad (1.5)$$

where δ is a fraction less than unity, or

$$u_o = K_1(u_a + u_b) + K_2(u_c + u_d), \quad (1.6)$$

where K_1 and K_2 are constant coefficients given by

$$\left. \begin{aligned} K_1 &= 1/2(\delta+1), & (a) \\ K_2 &= \delta/2(\delta+1). & (b) \end{aligned} \right\} (1.7)$$

When $b=c=d$, $a < b$, and $a/b = \delta$, we have

$$u_o = \frac{u_a}{(1+\delta)^2} + \frac{\delta u_o}{(1+\delta)^2} + \frac{\delta(u_b + u_d)}{2(1+\delta)}. \quad (1.8)$$

The process of adjustment which makes the values of a harmonic function agree with the four-point influence equation at each network point may be called a process of harmonization. The most commonly used method for harmonization is that of iteration¹ in which, arbitrarily assumed or judiciously selected, initial values are successively improved until practical convergence is obtained. Another procedure which has proved effective is that of relaxation.² More recently harmonization in rectangular areas has been effected by a solution of a special system of difference equations.³

The object of this paper is to develop an analytical process of harmonization, which yields key values of the harmonic function U in rectangular figures in terms of certain simple combinations of the given boundary values. From the key values all, except two or three, of the remaining interior values can be calculated directly by the simple Liebmann formula (1.3). For two or three points (1.8) may be required. The present treatment is believed to result in simple derivations, a relatively large scope, and a material reduction in the labor required to effect a solution. The method is directly applicable to composite rectangular figures, such as angles, channels, T-figures, Z-figures, etc. The proposed process has been used in numerical cases and found to be rapid and accurate.

The rectangles treated have an arbitrary width m and length n' . The width is always represented by an integer. Specific solutions are obtained for rectangles $(3 \times n')$ and $(4 \times n')$ for all positive values of n' between 2 and 14. The results are summarized in tables at the end of the paper. The tables are applicable to any value of n greater than 14.

Key Values. Definition

A set of interior values of U , so located in the region R to which the function U applies that, from these values, the values of U at all other network points can be calculated by the Liebmann or four-point formulas, applied either normally or diagonally, is defined as a set of key values. In a square (3×3) , Fig. 2(b), any one of the four interior values u_1 , u_2 , u_3 , and u_4 is a key value for the whole region. Thus if u_1 be

4. KEY VALUES FOR INCOMPLETE AND FULL (4×3) RECTANGLES, FIG. 4

(a) A rectangle $4(2+\delta)$, Fig. 4(a), will be spoken of as an incomplete (4×3) rectangle.

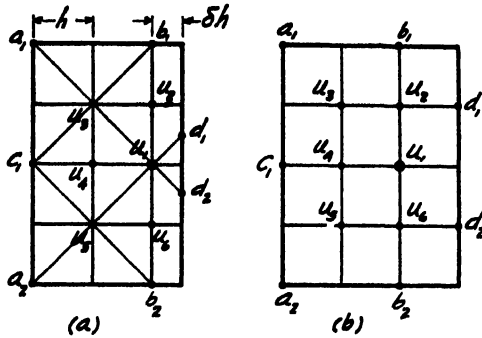


FIG. 4.

To find the key value u_1 we again use (1.6) and (1.3) and obtain the following three equations:

$$\left. \begin{aligned} u_1 &= K_1(d_1+d_2)+K_2(u_8+u_9), & (a) \\ 4u_8 &= u_1+a_1+b_1+c_1, & (b) \\ 4u_9 &= u_1+a_2+b_2+c_1. & (c) \end{aligned} \right\} (4.1)$$

Upon solution we have

$$\left(\frac{4+3\delta}{\delta} \right) u_1 = \frac{2}{\delta} (d_1+d_2) + \frac{1}{2} (a_1+a_2+b_1+b_2) + c_1. \quad (4.2)$$

(b) Putting δ equal to unity in (4.2) we obtain

$$7u_1 = \frac{1}{2} (a_1+a_2+b_1+b_2) + 2(d_1+d_2) + c_1. \quad (4.3)$$

This is the key value for a full (4×3) rectangle, Fig. 4(b), which agrees with (3.3) giving the key value for a full (3×4) rectangle at the same point.

5. KEY VALUES FOR INCOMPLETE AND FULL (4×4) SQUARE, FIG. 5

(a) A rectangle $4(3+\delta)$ will be spoken of as an incomplete (4×4) square, Fig. 5(a). There are three ways to obtain an expression for the key value u_1 . The simplest procedure is to express u_1 in terms of u_2, u_3, u_4 , and u_5 and then to express these quantities in terms of u_1 . Using (1.3) and (1.6) we obtain the following five expressions:

$$\left. \begin{aligned} 4u_1+u_2+u_3+u_4+u_5, & & (a) \\ 4u_2=u_1+a_2+c_2+c_3, & & (b) \\ 4u_4=u_1+a_3+c_3+c_4, & & (c) \\ u_2=K_1(a_1+c_1')+K_2(u_1+c_2), & & (d) \\ u_5=K_1(a_4+c_1'')+K_2(u_1+c_4). & & (e) \end{aligned} \right\} (5.1)$$

Upon solution we have

$$\left(\frac{7+5\delta}{1+\delta} \right) u_1 = \frac{1}{2} (a_2+a_3+c_2+2c_3+c_4) + \frac{1}{(1+\delta)} (a_1+a_4+c_1'+c_1'') + \frac{\delta}{(1+\delta)} (c_3+c_4). \quad (5.2)$$

(b) Putting δ equal unity and observing that then $c_1'=c_1''=c_1$ we obtain

$$6u_1 = \frac{1}{2} \sum_{i=1}^4 a_i + \sum_{i=1}^4 c_i. \quad (5.3)$$

This gives the key value for a full (4×4) square.

(c) Equation (5.3) also yields the key value for a rectangle (4×3) . Putting δ equal zero, changing the notation of the boundary values to

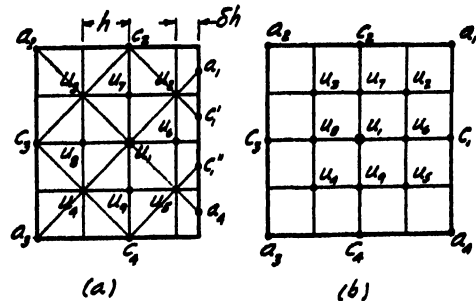


FIG. 5.

agree with that of Fig. 4(b), and observing that when $\delta=0$, $c_1'=a_1=d_1$, and $c_1''=a_4=d_2$, we obtain (4.3) for a full (4×3) rectangle.

6. LAPLACIAN PERIMETERS

In the further development of the method under consideration, extensive use is made of the key values for the basic figures treated in the preceding sections. By combining the results for the elementary rectangles, solutions are built up for longer rectangles. It is therefore convenient to simplify the forms of the expressions of the

key values already derived. Specifically we shall write each key value in the form

$$\left. \begin{aligned} \alpha u &= P + u', & (a) \\ \alpha u &= P + u' + u'', & (b) \end{aligned} \right\} (6.1)$$

or,

in which α is a constant, P is a quantity determined by the known boundary values, and u' , u , u'' are consecutive key values. The quantity P in (6.1) will, for lack of a better term, be referred to as a Laplacian perimeter or more briefly perimeter. Since we have eight basic rectangles, we shall need eight Laplacian perimeters: four for the incomplete rectangles and four for the full figures. We shall denote these perimeters by the generic symbols p and P , for the incomplete and full rectangles, respectively. As already stated, these perimeters are to contain only known boundary values. It will be seen later that, in each configuration, one or at the most two, boundary values will be characteristically missing in comparison with the full expression for the key value. Thus, incomplete and full (3×3) squares will always be used in such a manner that c_1 will be unknown, and in the full (3×4) rectangle both b_1 and b_2 or at least b_1 , will be unknown. In the definitions of the eight Laplacian perimeters that follow, we are guided entirely by such considerations.

7. DEFINITIONS OF LAPLACIAN PERIMETERS

Notation: The notation $P_{m \times n}$ or $(P)_{m \times n}$ denotes a Laplacian perimeter in a rectangle $(m \times n)$.

Full rectangles— P	Incomplete rectangles— p
$P_{3 \times 3} = \bar{P}$,	$p_{3 \times 3} = \bar{p}$,
$P_{3 \times 4} = P''$,	$p_{3 \times 4} = p''$,
$P_{4 \times 3} = P'$,	$p_{4 \times 3} = p'$,
$P_{4 \times 4} = P$,	$p_{4 \times 4} = p$.

Definitions:

1. Incomplete (3×3) square, Fig. 2(a),

$$\bar{p} = \frac{4}{\delta}(a_1 + a_2) + 4a_3 + c_1 + c_2. \quad (7.1)$$

2. Full (3×3) square, Fig. 2(b),

$$\bar{P} = 4 \sum_{i=1}^3 a_i + c_1 + c_2. \quad (7.2)$$

3. Incomplete (3×4) rectangle, Fig. 3(a),

$$p'' = a_1 + c_1 + 4(d_1 + d_2) + \left(\frac{2}{1 + \delta} \right) (a_2 + b_2 + \delta c_1). \quad (7.3)$$

4. Full (3×4) rectangle, Fig. 3(b),

$$P'' = (a_1 + a_2 + 2c_1) + 4(d_1 + d_2) + \text{known } b\text{'s}. \quad (7.4)$$

5. Incomplete (4×3) rectangle, Fig. 4(a),

$$p' = \frac{1}{2}(a_1 + a_2 + b_1 + b_2) + \frac{2}{\delta}(d_1 + d_2). \quad (7.5)$$

6. Full (4×3) rectangle, Fig. 4(b),

$$P' = \frac{1}{2}(a_1 + a_2 + b_1 + b_2) + 2(d_1 + d_2). \quad (7.6)$$

7. Incomplete (4×4) square, Fig. 5(a),

$$p = \frac{1}{2}(a_2 + a_3 + c_3 + c_4) + \left(\frac{1}{1 + \delta} \right) (a_1 + a_4 + c_1' + c_1'') + \frac{\delta(c_2 + c_4)}{1 + \delta}. \quad (7.7)$$

8. Full (4×4) square, Fig. 5(b),

$$P = \frac{1}{2} \sum_{i=1}^4 a_i + \sum \text{known } c\text{'s}. \quad (7.8)$$

8. RELATION BETWEEN $P_{m \times n}$ AND $P_{m \times n}'$

(a) In all figures treated in this paper there are as many Laplacian perimeters as there are key points, or key values.

(b) The numeral subscript in the perimeter refers to the key point of the same number. Thus P_i refers to key point (i) of a key value u_i .

(c) The notation $(P_i)_{m \times n}$ denotes a perimeter at a point (i) in a rectangle $(m \times n)$.

(d) When the length of a rectangular figure is increased by λ , only the last perimeter P_n in the rectangle $(4 \times n)$ changes. In rectangles $(3 \times n)$ and $(4 \times n)$ the relation between the old and

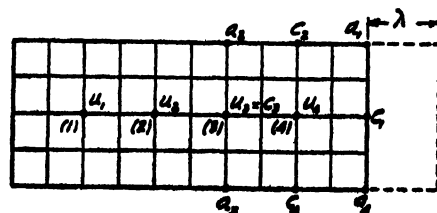


FIG. 6.

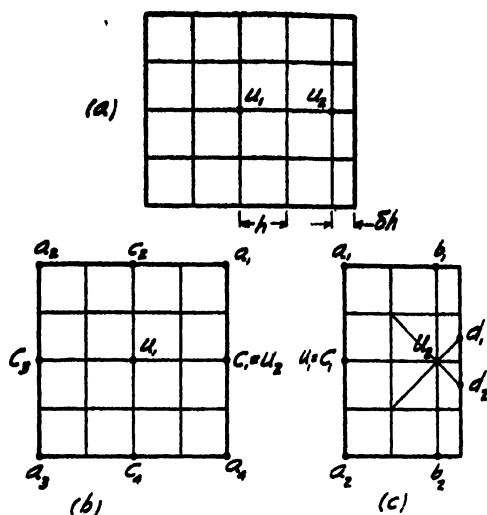


FIG. 7.

new perimeters is

$$(P_k)_{m \times n} = (P_k)_{m \times n'} + u_{k+1}, \quad (8.1)$$

k referring to the figure ($m \times n$).

Consider for concreteness the rectangle (4×10), Fig. 6. In the last square having its center at point (4), the Laplacian perimeter $(P_4)_{4 \times 10}$ is by definition, (7.8),

$$(P_4)_{4 \times 10} = \frac{1}{2} \sum_{i=1}^4 a_i + c_1 + c_2 + c_4,$$

since $c_3 = u_3$ is unknown. If the length $n = 10$ is increased by λ

$$(P_4)_{4(10+\lambda)} = \frac{1}{2} \sum_{i=1}^4 a_i + c_2 + c_4,$$

since now both $c_1 = u_5$, as well as $c_3 = u_3$, are unknown, so that

$$(P_4)_{4 \times 10} = (P_4)_{4(10+\lambda)} + u_5.$$

We note, however, that P_1 , P_2 , and P_3 remain unchanged, or in general

$$(P_i)_{m \times n} = (P_i)_{m \times n'}, \quad \text{for } i < k. \quad (8.2)$$

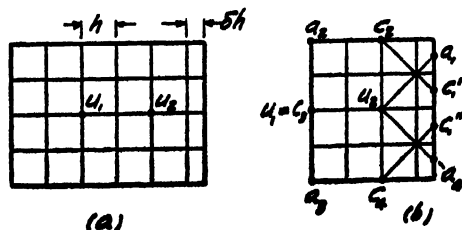


FIG. 8.

9. BASIC KEY VALUES IN TERMS OF P AND p

1. Incomplete (3×3) square, Fig. 2(a). Eq. (2.2) becomes

$$\begin{aligned} \alpha u_1 &= p + c_1, & (a) \\ \alpha &= (8 + 7\delta)/\delta. & (b) \end{aligned} \quad (9.1)$$

2. Full (3×3) square, Fig. 2(b). Eq. (2.3) becomes

$$15u_1 = P + c_1. \quad (9.2)$$

3. Incomplete (3×4) rectangle, Fig. 3(a). Eq. (3.2) becomes

$$\begin{aligned} \alpha'' u_1 &= p'' + b_1, & (a) \\ \alpha'' &= (15 + 13\delta)/(1 + \delta). & (b) \end{aligned} \quad (9.3)$$

4. Full (3×4) rectangle, Fig. 3(b). Eq. (3.3) becomes

$$14u_1 = P'' + \sum \text{unknown } b\text{'s}. \quad (9.4)$$

5. Incomplete (4×3) rectangle, Fig. 4(a). Eq. (4.2) becomes

$$\begin{aligned} \alpha' u &= p' + c_1, & (a) \\ \alpha' &= (4 + 3\delta)/\delta. & (b) \end{aligned} \quad (9.5)$$

6. Full (4×3) rectangle, Fig. 4(b). Eq. (4.3) becomes

$$7u_1 = P' + c_1. \quad (9.6)$$

7. Incomplete (4×4) square, Fig. 5(a). Eq. (5.2) becomes

$$\begin{aligned} \alpha u_1 &= p + c_3, & (a) \\ \alpha &= (7 + 5\delta)/(1 + \delta). & (b) \end{aligned} \quad (9.7)$$

8. Full (4×4) square, Fig. 5(b). Eq. (5.3) becomes

$$6u_1 = P + \sum \text{unknown } c\text{'s}. \quad (9.8)$$

10. GENERAL METHOD

By combining the solutions for the values of U' for the basic cases it is now possible to obtain *continuous* solutions for the key values in rectangles ($m \times n'$), where n' is any positive number greater than 2 and $m = 3, 4$.

The general method rests on the following simple proposition: If a solution for the last key value u_k is available for a rectangle ($m \times n$), n any positive integer, it is possible to extend this solution to u_{k+1} in rectangles $m(n + \delta)$, or $m(n + 1 + \delta)$. The range of δ is $0 \leq \delta \leq 1$. As a matter of convenience it is best to start with n equal to an even positive integer.

Any rectangle ($m \times n'$) may be viewed as a rectangle ($m \times n$), n the nearest even integer, overlapped by an incomplete ($m \times 3$) or ($m \times 4$). Thus, a rectangle $4(6+\delta)$ may be looked upon as a (4×6) overlapped by an incomplete (4×3), and a rectangle $4(7+\delta)$ may similarly be viewed as a (4×6) overlapped by an incomplete (4×4), Figs. 7 and 8. Now, in the rectangle ($m \times n$) the last key value u_k may be written as

$$u_k = C_1 P_1 + C_2 P_2 + \cdots C_k P_k \quad (10.1)$$

in which the C 's are constant coefficients and the P 's are perimeters for the rectangle ($m \times n$). Remembering the relation between $P_{m \times n}$ and $P_{m \times n'}$, (8.1, 8.2) we have

$$u_k = C_1 P_1 + C_2 P_2 + \cdots C_k P_k + C_k u_{k+1}, \quad (10.2)$$

in which the C 's are the same as (10.1) but the P 's now refer to the rectangle ($m \times n'$). Also from the incomplete rectangles ($m \times 3$) or ($m \times 4$), Section 9, we have

$$\alpha u_{k+1} = p_{k+1} + u_k. \quad (10.3)$$

Solving (10.2) and (10.3) we obtain

$$u_{k+1} = (\alpha - C_k)^{-1} (C_1 P_1 + C_2 P_2 + \cdots C_k P_k) + (\alpha - C_k)^{-1} p_{k+1}. \quad (10.4)$$

We thus have a general relation between the coefficients for the last key value u_k in the rectangle ($m \times n$) and the coefficients for the next key value u_{k+1} in the rectangle $m(n+\delta)$, or $m(n+1+\delta)$. Specifically

$$\left. \begin{aligned} C_{k+1} &= (\alpha - C_k)^{-1}, & (a) \\ \text{and for all other coefficients} & & \\ (C_i)_{m \times n'} &= (C_i)_{m \times n} (\alpha - C_k)^{-1}. & (b) \end{aligned} \right\} (10.5)$$

In using these formulas, it must be remembered that there are four α 's, which depend on whether $m=3$ or $m=4$, and on whether n' differs from n by less than unity or by more than unity. Thus, for rectangles $4(n+1+\delta)$ and $4(n+\delta)$ we use α and α' , respectively, which are given by (9.7b) and (9.5b). Similarly for rectangles $3(n+1+\delta)$ and $3(n+\delta)$ we use α'' and $\bar{\alpha}$, respectively, given by (9.3b) and (9.1b).

The determination of u_{k+1} reduces the rectangle ($m \times n'$) to that of a rectangle ($m \times n$) for which a solution for all key values is assumed to be known.

Hence, starting with the expressions for the key value u_1 in a (4×4) square, (9.8), one can successively write down the formulas for the last key values in all subsequent rectangles ($4 \times n'$). Similarly, starting with the expression for u_1 in the rectangle (3×4), (9.4), we can write the formulas for all rectangles ($3 \times n'$).

11. SUCCESSIVE FACTORS $(\alpha - C_k)^{-1}$

For brevity we simplify our notation and write

$$(\alpha - C_k)^{-1} = F_i, \quad (11.1)$$

and

$$(F, C, \alpha)_{m(n+\delta)} = (F, C, \alpha)_{n+\delta}. \quad (11.2)$$

We then have

$$F_{4+\delta} = (C_2)_{4+\delta} = [\alpha_{2+\delta} - (C_1)_4]^{-1}, \quad (11.3)$$

$$F_{5+\delta} = (C_2)_{5+\delta} = [\alpha_{3+\delta} - (C_1)_4]^{-1}, \quad (11.4)$$

$$F_{6+\delta} = (C_2)_{6+\delta} = [\alpha_{2+\delta} - (C_2)_6]^{-1} = (\alpha_{2+\delta} - F_6)^{-1}, \quad (11.5)$$

$$F_{7+\delta} = (C_2)_{7+\delta} = [\alpha_{3+\delta} - (C_2)_6]^{-1} = (\alpha_{3+\delta} - F_6)^{-1}, \quad (11.6)$$

$$F_{8+\delta} = (C_2)_{8+\delta} = [\alpha_{2+\delta} - (C_2)_8]^{-1} = (\alpha_{2+\delta} - F_8)^{-1}. \quad (11.7)$$

In general, for n an even integer

$$\left. \begin{aligned} F_{n+\delta} &= (\alpha_{2+\delta} - F_n)^{-1}, & (a) \\ F_{n+1+\delta} &= (\alpha_{3+\delta} - F_n)^{-1}. & (b) \end{aligned} \right\} (11.8)$$

For example, to find $F_{4(5+\delta)}$ we write

$$F_{4(5+\delta)} = (\alpha_{4(3+\delta)} - F_{4 \times 4})^{-1} = [\alpha_{4(3+\delta)} - (C_1)_{4 \times 4}]^{-1}. \quad (11.9)$$

Substituting

$$\alpha_{4(3+\delta)} = (7+5\delta)/(1+\delta), \quad (9.7b)$$

and

$$(C_1)_{4 \times 4} = \frac{1}{2}, \quad (9.8)$$

we have

$$F_{4(5+\delta)} = (41+29\delta)/(1+\delta)6. \quad (11.9)$$

We can now write down the expression for $(u_2)_{4(5+\delta)}$. Thus by (10.4)

$$(u_2)_{4(5+\delta)} = F_{4(5+\delta)} (C_1)_{4 \times 4} P_1 + F_{4(5+\delta)} p_2. \quad (11.10)$$

Calculating several values of F it is found that these factors converge rapidly, and that

TABLE I.

n'	C_1	C_2	C_3	C_4	C_5	C_6	C_7
$4+\delta$	$(C_1)_4 F_{4+\delta}$	$F_{4+\delta}$					
$5+\delta$	$(C_1)_4 F_{5+\delta}$	$F_{5+\delta}$					
$6+\delta$	$(C_1)_4 F_6 F_{6+\delta}$	$F_6 F_{6+\delta}$	$F_{6+\delta}$				
$7+\delta$	$(C_1)_4 F_6 F_7 F_{7+\delta}$	$F_6 F_7 F_{7+\delta}$	$F_7 F_{7+\delta}$				
$8+\delta$	$(C_1)_4 F_6 F_8 F_{8+\delta}$	$F_6 F_8 F_{8+\delta}$	$F_8 F_{8+\delta}$	$F_{8+\delta}$			
$9+\delta$	$(C_1)_4 F_6 F_8 F_9 F_{9+\delta}$	$F_6 F_8 F_9 F_{9+\delta}$	$F_8 F_9 F_{9+\delta}$	$F_9 F_{9+\delta}$			
$10+\delta$	$(C_1)_4 F_6 F_8 F_{10} F_{10+\delta}$	$F_6 F_8 F_{10} F_{10+\delta}$	$F_8 F_{10} F_{10+\delta}$	$F_{10} F_{10+\delta}$	$F_{10+\delta}$		
$11+\delta$	$(C_1)_4 F_6 F_8 F_{10} F_{11+\delta}$	$F_6 F_8 F_{10} F_{11+\delta}$	$F_8 F_{10} F_{11+\delta}$	$F_{10} F_{11+\delta}$	$F_{11+\delta}$		
$12+\delta$	$(C_1)_4 F_6 F_8 F_{10} F_{12} F_{12+\delta}$	$F_6 F_8 F_{10} F_{12} F_{12+\delta}$	$F_8 F_{10} F_{12} F_{12+\delta}$	$F_{10} F_{12} F_{12+\delta}$	$F_{12} F_{12+\delta}$	$F_{12+\delta}$	
$13+\delta$	$(C_1)_4 F_6 F_8 F_{10} F_{12} F_{13+\delta}$	$F_6 F_8 F_{10} F_{12} F_{13+\delta}$	$F_8 F_{10} F_{12} F_{13+\delta}$	$F_{10} F_{12} F_{13+\delta}$	$F_{12} F_{13+\delta}$	$F_{13+\delta}$	
$14+\delta$	$(C_1)_4 F_6 F_8 F_{10} F_{12} F_{14} F_{14+\delta}$	$F_6 F_8 F_{10} F_{12} F_{14} F_{14+\delta}$	$F_8 F_{10} F_{12} F_{14} F_{14+\delta}$	$F_{10} F_{12} F_{14} F_{14+\delta}$	$F_{12} F_{14} F_{14+\delta}$	$F_{14} F_{14+\delta}$	$F_{14+\delta}$
$15+\delta$	$(C_1)_4 F_6 F_8 F_{10} F_{12} F_{14} F_{15+\delta}$	$F_6 F_8 F_{10} F_{12} F_{14} F_{15+\delta}$	$F_8 F_{10} F_{12} F_{14} F_{15+\delta}$	$F_{10} F_{12} F_{14} F_{15+\delta}$	$F_{12} F_{14} F_{15+\delta}$	$F_{14} F_{15+\delta}$	$F_{15+\delta}$

they become essentially constant at $n=6$ when $m=4$, and at $n=4$ when $m=3$. Thus, to five decimal places,

$$\left. \begin{aligned} F_{4(n+\delta)} &= F_{4(7+\delta)}, & n \text{ odd}; & \quad (a) \\ F_{4(n+\delta)} &= F_{4(8+\delta)}, & n \text{ even}; & \quad (b) \\ F_{8(n+\delta)} &= F_{8(5+\delta)}, & n \text{ odd}; & \quad (c) \\ F_{8(n+\delta)} &= F_{8(6+\delta)}, & n \text{ even}. & \quad (d) \end{aligned} \right\} (11.11)$$

It is easy to verify (11.11) for integral values of n , i.e., $\delta=0$. For example

$$\begin{aligned} F_{4 \times 15} - F_{4 \times 7} &= .0000031, \\ F_{4 \times 16} - F_{4 \times 8} &= .0000042, \\ F_{8 \times 15} - F_{8 \times 5} &= .00000165, \\ F_{8 \times 16} - F_{8 \times 6} &= .00000189. \end{aligned}$$

But by (11.8) for n an even or odd integer

$$F_{n+2+\delta} - F_{n+\delta} = (\alpha - F_{n+2})^{-1} - (\alpha - F_n)^{-1}.$$

Since the α 's are the same, and for integral values of n , $F_{n+2} = F_n$, for $n > 6$, $m=4$, and $n > 4$, $m=3$ it follows that the relations given in (11.11) are correct to five decimal places.

12. TABLE OF MASTER COEFFICIENTS FOR THE LAST KEY VALUE u_{k+1}

From (10.5) it is possible to write down all coefficients from C_1 to C_{k+1} for the last key value in all rectangles $(m \times n')$. Thus using the abridged notation from the preceding section and the general relation

$$(C_i)_{n'} = (C_i)_n F_{n'}, \quad (10.5b)$$

we have

$$\left. \begin{aligned} (C_1)_{4+\delta} &= (C_1)_4 F_{4+\delta}, & (a) \\ (C_1)_{5+\delta} &= (C_1)_4 F_{5+\delta}, & (b) \\ (C_1)_{6+\delta} &= (C_1)_6 F_{6+\delta} = (C_1)_4 F_6 F_{6+\delta}, & (c) \\ (C_1)_{7+\delta} &= (C_1)_6 F_{7+\delta} = (C_1)_4 F_6 F_7 F_{7+\delta}, \text{ etc.} & (d) \end{aligned} \right\} (12.1)$$

Similarly,

$$\left. \begin{aligned} (C_2)_{4+\delta} &= F_{4+\delta}, & (a) \\ (C_2)_{5+\delta} &= F_{5+\delta}, & (b) \\ (C_2)_{6+\delta} &= (C_2)_6 F_{6+\delta} = F_6 F_{6+\delta}, & (c) \\ (C_2)_{7+\delta} &= (C_2)_6 F_{7+\delta} = F_6 F_7 F_{7+\delta}, \text{ etc.} & (d) \end{aligned} \right\} (12.2)$$

A summary of all coefficients for the last key value in rectangle $(m \times n')$, $m=3, 4$, $4 \leq n \leq 16$ is given in Table I. Hereafter the last key value will be denoted by u_k in all rectangles.

13. GENERAL RELATIONS BETWEEN THE COEFFICIENTS IN TABLE I

Inspection of Table I shows three important relations between the coefficients for the last key value.

(a) For any fixed value n'

$$\begin{aligned} (C_1)_{n'} &= (C_1)_4 F_6 F_8 F_{10} \cdots F_{n'}; & C_2 &= C_1 / (C_1)_4; \\ C_3 &= C_2 / F_6; & C_4 &= C_2 / F_8; \\ C_i &= C_{i-1} / F_{2i}; & C_k &= F_{n'}. \end{aligned} \quad (13.1)$$

(b) The last four coefficients $C_k, C_{k-1}, C_{k-2}, C_{k-3}$ become constant for all even values of n , and for all odd values of n , at $n \geq 12$. Thus in rectangles $m(n+\delta)$ for n even,

$$C_k = F_{n+\delta}; \quad C_{k-1} = F_n F_{n+\delta}; \quad (13.2)$$

$$C_{k-2} = F_{n-2} F_n F_{n+\delta}, \quad C_{k-3} = F_{n-4} F_{n-2} F_n F_{n+\delta}.$$

For n odd,

$$C_k = F_{n+\delta}; \quad C_{k-1} = F_{n-1} F_{n+\delta};$$

$$C_{k-2} = F_{n-3} F_{n-1} F_{n+\delta}; \quad (13.3)$$

$$C_{k-3} = F_{n-5} F_{n-3} F_{n-1} F_{n+\delta}.$$

Introducing the approximations of (11.1) we have for n even,

$$C_k = F_{8+\delta}; \quad C_{k-1} = F_8 F_{8+\delta}; \quad (13.4)$$

$$C_{k-2} = F_8^2 F_{8+\delta}; \quad C_{k-3} = F_8^3 F_{8+\delta}.$$

For n odd,

$$C_k = F_{7+\delta}; \quad C_{k-1} = F_8 F_{7+\delta}; \quad (13.5)$$

$$C_{k-2} = F_8^2 F_{7+\delta}; \quad C_{k-3} = F_8^3 F_{7+\delta}.$$

(c) For n greater than six, assuming $F_8 = F_{10} = F_{12}$, etc.

$$(C_i)_{n'+\delta} = (C_i)_n F_8. \quad (13.6)$$

14. MAIN TABLES AND THEIR USE

From the expressions given in Table I, and the simplifications for F made in Section 11, it is possible to calculate the coefficients $C_1, C_2, \dots, C_k$ for the 1st key value in a rectangle of arbitrary length n' . Such calculations were made for $m(n+\delta)$, $m=3, 4$; $n=2, 3, 4, 5, 6, \dots, 13$, and $\delta=.1, .2, .3, \dots, .8, .9, 1.0$. The results are given in Tables II and III.

These tables are sufficient to determine all key values in rectangles ($m \times n'$), for values of n' between 2 and infinity. The procedure is as follows.

First we calculate the last key value u_n using the appropriate coefficients in the tables. We can then treat the rectangle as if it were shorter and its length n was an even integer. Next we find u_{k-1} and then view the rectangle as if it were further shortened by 2 units. We thus can find u_{k-2}, u_{k-3} , etc. and continue the process until all key values are found.

It should be observed that in calculating u_{k-1} the last perimeter P_{k-1} must be modified by the addition of u_k . In general, in calculating u , we add u_{i+1} to P_i . Thus

$$u_i = C_1 P_1 + C_2 P_2 + \dots + C_i (P_i + u_{i+1}). \quad (14.1)$$

TABLE II. $3(n+\delta)$.

n	$\delta \rightarrow$	$\delta \rightarrow$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
u_1	2	C_1	.01149	.02128	.02970	.03704	.04348	.04918	.05426	.05882	.06294	.06667
u_1	3	C_1	.06748	.06818	.06878	.06931	.06977	.07018	.07054	.07087	.07116	.07143
u_2	4	C_1	.00082	.00152	.00213	.00265	.00312	.00353	.00389	.00422	.00452	.00478
		C_2	.01150	.02131	.02977	.03714	.04361	.04935	.05447	.05907	.06322	.06699
u_2	5	C_1	.00484	.00490	.00494	.00498	.00501	.00504	.00506	.00509	.00511	.00513
		C_2	.06781	.06852	.06912	.06965	.07012	.07053	.07090	.07123	.07152	.07179
u_3	6	C_1	.00006	.00011	.00015	.00019	.00022	.00025	.00028	.00030	.00032	.00034
		C_2	.00083	.00153	.00214	.00267	.00313	.00354	.00391	.00424	.00454	.00481
		C_3	.01150	.02131	.02977	.03714	.04361	.04935	.05448	.05907	.06322	.06699
u_3	7	C_1	.00035	.00035	.00035	.00036	.00036	.00036	.00036	.00037	.00037	.00037
		C_2	.00487	.00492	.00496	.00500	.00503	.00506	.00509	.00511	.00514	.00515
		C_3	.06781	.06852	.06912	.06965	.07012	.07053	.07090	.07123	.07153	.07180
u_4	8	C_1	.00000	.00001	.00001	.00001	.00002	.00002	.00002	.00002	.00002	.00002
		C_2	.00006	.00011	.00015	.00019	.00022	.00025	.00028	.00030	.00033	.00035
		C_3	.00083	.00153	.00214	.00267	.00313	.00354	.00391	.00424	.00454	.00481
		C_4	.01150	.02131	.02977	.03714	.04361	.04935	.05448	.05907	.06322	.06699
u_4	9	C_1	.00002	.00003	.00003	.00003	.00003	.00003	.00003	.00003	.00003	.00003
		C_2	.00035	.00035	.00035	.00036	.00036	.00036	.00036	.00037	.00037	.00037
		C_3	.00487	.00492	.00496	.00500	.00503	.00506	.00509	.00511	.00514	.00515
		C_4	.06781	.06852	.06912	.06965	.07012	.07053	.07090	.07123	.07153	.07180

TABLE III. $4(n+\delta)$.

n	δ	C_1	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
u_1	2	C_1	.02326	.04348	.06122	.07692	.09091	.10345	.11475	.12500	.13433	.14286
u_1	3	C_1	.14667	.15000	.15294	.15556	.15789	.16000	.16190	.16364	.16522	.16667
u_2	4	C_1	.00389	.00730	.01031	.01299	.01538	.01754	.01950	.02128	.02290	.02439
		C_2	.02335	.04380	.06186	.07792	.09231	.10526	.11699	.12766	.13740	.14634
u_2	5	C_1	.02506	.02564	.02616	.02662	.02703	.02740	.02773	.02804	.02832	.02857
		C_2	.15034	.15385	.15694	.15970	.16216	.16438	.16639	.16822	.16990	.17143
u_3	6	C_1	.00067	.00125	.00177	.00223	.00264	.00301	.00334	.00365	.00393	.00418
		C_2	.00400	.00751	.01061	.01336	.01583	.01805	.02007	.02190	.02357	.02510
		C_3	.02335	.04380	.06187	.07795	.09235	.10532	.11706	.12774	.13749	.14644
u_3	7	C_1	.00430	.00440	.00449	.00457	.00464	.00470	.00476	.00481	.00486	.00490
		C_2	.02579	.02639	.02692	.02740	.02782	.02820	.02855	.02886	.02915	.02941
		C_3	.15045	.15396	.15706	.15982	.16229	.16451	.16653	.16836	.17003	.17157
u_4	8	C_1	.00011	.00021	.00030	.00038	.00045	.00052	.00057	.00063	.00067	.00072
		C_2	.00069	.00129	.00182	.00229	.00272	.00310	.00344	.00376	.00404	.00431
		C_3	.00401	.00752	.01062	.01337	.01584	.01807	.02008	.02192	.02359	.02513
		C_4	.02335	.04381	.06187	.07795	.09235	.10532	.11706	.12774	.13750	.14645
u_4	9	C_1	.00074	.00075	.00077	.00078	.00080	.00081	.00082	.00083	.00083	.00084
		C_2	.00443	.00453	.00462	.00470	.00477	.00484	.00490	.00495	.00500	.00505
		C_3	.02581	.02642	.02695	.02742	.02784	.02823	.02857	.02889	.02917	.02944
		C_4	.15045	.15396	.15706	.15982	.16229	.16452	.16653	.16836	.17004	.17157
u_5	10	C_1	.00002	.00004	.00005	.00007	.00008	.00009	.00010	.00011	.00012	.00012
		C_2	.00012	.00022	.00031	.00039	.00047	.00053	.00059	.00064	.00069	.00074
		C_3	.00069	.00129	.00182	.00229	.00272	.00310	.00345	.00376	.00405	.00431
		C_4	.00401	.00752	.01062	.01337	.01584	.01807	.02008	.02192	.02359	.02513
		C_5	.02335	.04381	.06187	.07795	.09235	.10532	.11706	.12774	.13750	.14645
u_5	11	C_1	.00013	.00013	.00013	.00013	.00014	.00014	.00014	.00014	.00014	.00014
		C_2	.00076	.00078	.00079	.00081	.00082	.00083	.00084	.00085	.00086	.00087
		C_3	.00443	.00453	.00462	.00470	.00478	.00484	.00490	.00496	.00501	.00505
		C_4	.02581	.02642	.02695	.02742	.02784	.02823	.02857	.02889	.02917	.02944
		C_5	.15045	.15396	.15706	.15982	.16229	.16452	.16653	.16836	.17004	.17157
u_5	12	C_1	.00000	.00001	.00001	.00001	.00001	.00002	.00002	.00002	.00002	.00002
		C_2	.00002	.00004	.00005	.00007	.00008	.00009	.00010	.00011	.00012	.00013
		C_3	.00012	.00022	.00031	.00039	.00047	.00053	.00059	.00065	.00069	.00074
		C_4	.00069	.00129	.00182	.00229	.00272	.00310	.00345	.00376	.00405	.00431
		C_5	.00401	.00752	.01062	.01337	.01584	.01807	.02008	.02192	.02359	.02513
		C_6	.02335	.04381	.06187	.07795	.09235	.10532	.11706	.12774	.13750	.14645
u_5	13	C_1	.00002	.00002	.00002	.00002	.00002	.00002	.00002	.00002	.00002	.00002
		C_2	.00013	.00013	.00013	.00014	.00014	.00014	.00014	.00015	.00015	.00015
		C_3	.00076	.00078	.00079	.00081	.00082	.00083	.00084	.00085	.00086	.00087
		C_4	.00443	.00453	.00462	.00470	.00478	.00484	.00490	.00496	.00501	.00505
		C_5	.02581	.02642	.02695	.02742	.02784	.02823	.02857	.02889	.02917	.02944
		C_6	.15045	.15396	.15706	.15982	.16229	.16452	.16653	.16836	.17004	.17157

Inspection of Tables II and III shows that the coefficients for u_k rapidly approach convergence. In rectangles ($3 \times n'$), Table II, convergence is reached at $n=9$. For values of n greater than 9, C_k , C_{k-1} , C_{k-2} , C_{k-3} are the same as C_4 , C_3 , C_2 , and C_1 , respectively, at $n=9$ or $n=8$ depending on whether n is odd or even.

In rectangles ($4 \times n'$) convergence is reached at $n=13$. The coefficients for values of δ other

than those given in Tables II and III can be found by interpolation.

15. AUXILIARY TABLES

It is sometimes desirable to be able to find directly a value of $u_i \neq u_k$ without having to find first u_k , $u_{k-1} \dots u_{i+1}$. This is especially useful in very long rectangles. This can be done by means of Tables IV and V, which give the coefficients

for any $u_i \neq u_k$ independent of all other key values. Tables IV and V are for regular rectangles, with integral lengths. If the length is not an integer it is necessary first to find u_k .

Another advantage of Tables IV and V lies in the fact that all key values are given in terms of the one set of P 's calculated for the rectangle ($m \times n$), and no modification of these are needed.

Inspection of Tables IV and V shows that when n equals 14 the coefficients have become stabilized and that these coefficients may be used for all subsequent values of n . Specifically, for any integral value of n greater than 14.

(a) The coefficients of u_1 , u_2 , and u_3 are the same as of u_1 , u_2 , and u_3 for $n = 14$.

(b) For n even, the coefficients C_{k-1} , C_{k-2} , etc. of u_{k-1} and u_{k-2} are the same as of u_5 and u_4 , respectively, at $n = 14$. For n odd, the corre-

sponding coefficients are the same as of u_5 and u_4 , respectively, at $n = 13$.

(c) The coefficients of u_i , ($3 < i < k-2$), are constant, and essentially the same as for u_3 in $n = 14$. Thus, to four decimal places, n odd or even,

$$\left. \begin{aligned} (u_1)_{3 \times n} &= .0722P_1 + .0052(P_{-1} + P_{+1}) \\ &\quad + .0004(P_{-2} + P_{+2}), \quad (a) \\ (u_1)_{4 \times n} &= .1768P_1 + .0303(P_{-1} + P_{+1}) \\ &\quad + .0052(P_{-2} + P_{+2}) \\ &\quad + .0009(P_{-3} + P_{+3}), \quad (b) \end{aligned} \right\} (15.1)$$

since the last perimeter in (15.1) is $P_{\pm 3}$ it follows that u_i is affected by a relatively small length of immediately adjacent boundaries. If we limit the accuracy of the coefficient to four

TABLE IV. ($3 \times n$).

n	u	C_1	C_2	C_3	C_4	C_5	C_6
5	u_1	.07177	.00478				
6	u_1	.07179	.00513				
7	u_1	.07180	.00515	.00034			
	u_2	.00515	.07214	.00481			
8	u_1	.07180	.00515	.00037			
	u_2	.00515	.07216	.00515			
9	u_1	.07180	.00515	.00037	.00002		
	u_2	.00515	.07217	.00518	.00035		
	u_3	.00037	.00518	.07214	.00481		
10	u_1	.07180	.00515	.00037	.00003		
	u_2	.00515	.07217	.00518	.00037		
	u_3	.00037	.00518	.07217	.00515		
11	u_1	.07180	.00515	.00037	.00003	.00000	
	u_2	.00515	.07217	.00518	.00037	.00002	
	u_3	.00037	.00518	.07217	.00518	.00035	
	u_4	.00003	.00037	.00518	.07214	.00481	
12	u_1	.07180	.00515	.00037	.00003	.00000	
	u_2	.00515	.07217	.00518	.00037	.00003	
	u_3	.00037	.00518	.07217	.00518	.00037	
	u_4	.00003	.00037	.00518	.07217	.00515	
13	u_1	.07180	.00515	.00037	.00003	.00000	.00000
	u_2	.00515	.07217	.00518	.00037	.00003	.00000
	u_3	.00037	.00518	.07217	.00518	.00037	.00002
	u_4	.00003	.00037	.00518	.07217	.00518	.00035
	u_5	.00000	.00003	.00037	.00518	.07214	.00481
14	u_1	.07180	.00515	.00037	.00003	.00000	.00000
	u_2	.00515	.07217	.00518	.00037	.00003	.00000
	u_3	.00037	.00518	.07217	.00518	.00037	.00003
	u_4	.00003	.00037	.00518	.07217	.00518	.00037
	u_5	.00000	.00003	.00037	.00518	.07217	.00515

TABLE V. ($4 \times n$).

n	u	C_1	C_2	C_3	C_4	C_5	C_6
5	u_1	.17073	.02439				
6	u_1	.17143	.02857				
7	u_1	.17155	.02929	.00418			
	u_2	.02029	.17573	.02510			
8	u_1	.17157	.02941	.00490			
	u_2	.02941	.17647	.02941			
9	u_1	.17157	.02943	.00503	.00072		
	u_2	.02943	.17660	.03015	.00431		
	u_3	.00503	.03015	.17588	.02513		
10	u_1	.17157	.02944	.00505	.00084		
	u_2	.02944	.17662	.03028	.00505		
	u_3	.00505	.03028	.17662	.02944		
11	u_1	.17157	.02944	.00505	.00086	.00012	
	u_2	.02944	.17662	.03030	.00517	.00074	
	u_3	.00505	.03030	.17675	.03018	.00431	
	u_4	.00086	.00517	.03018	.17588	.02513	
12	u_1	.17157	.02944	.00505	.00087	.00014	
	u_2	.02944	.17662	.03030	.00519	.00087	
	u_3	.00505	.03030	.17677	.03030	.00505	
	u_4	.00087	.00519	.03030	.17662	.02944	
13	u_1	.17157	.02944	.00505	.00087	.00015	.00002
	u_2	.02944	.17662	.03030	.00519	.00089	.00013
	u_3	.00505	.03030	.17677	.03032	.00518	.00074
	u_4	.00087	.00520	.03032	.17675	.03018	.00431
	u_5	.00015	.00089	.00518	.03018	.17589	.02513
14	u_1	.17157	.02944	.00505	.00087	.00015	.00002
	u_2	.02944	.17662	.03030	.00520	.00089	.00015
	u_3	.00505	.03030	.17677	.03033	.00520	.00087
	u_4	.00087	.00520	.03033	.17677	.03030	.00505
	u_5	.00015	.00089	.00520	.03030	.17662	.02944

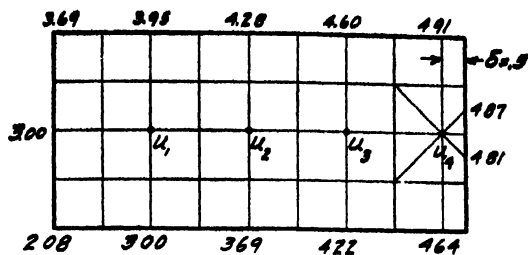


FIG. 9.

decimal places it is seen that in rectangles ($4 \times n$) the length of the influencing boundary does not extend beyond six lattice units measured from point (i) in either direction. In rectangles ($3 \times n$) the corresponding length is only four lattice units. These facts are useful in approximate solutions.

16. NUMERICAL EXAMPLE

Given a rectangle 4×8.5 with boundary values as shown in Fig. 9. Required to find all key values. Solution:

(a) We begin by calculating P_1 , P_2 , P_3 , and p_4 . The necessary Eqs. are (7.8) and (7.6). Thus

$$\begin{aligned} P_1 &= .5(4.28 + 3.69 + 2.08 + 3.69) \\ &\quad + 3.95 + 3.00 + 3.00 = 16.82, \\ P_2 &= .5(4.60 + 3.95 + 3.00 + 4.22) \\ &\quad + 4.28 + 3.69 = 15.86, \\ P_3 &= .5(4.91 + 4.28 + 3.69 + 4.64) \\ &\quad + 4.60 + 4.22 = 17.58, \\ p_4 &= .5(4.60 + 4.22 + 4.91 + 4.64) \\ &\quad + 4(4.87 + 4.81) = 47.91. \end{aligned}$$

We next find $u_k = u_4$. From Table III we have

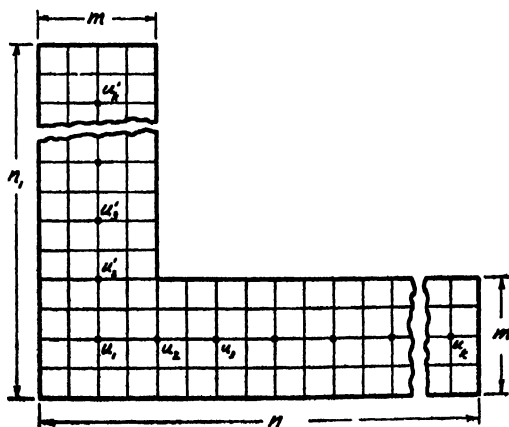


FIG. 10.

$$\begin{aligned} u_4 &= .00045P_1 + .00272P_2 + .01584P_3 \\ &\quad + .09235p_4 = 4.754. \end{aligned}$$

We now view the rectangle as a (4×8) and find u_3 . Thus

$$\begin{aligned} u_3 &= .00490P_1 + .02941P_2 \\ &\quad + .17157(P_3 + u_4) = 4.381. \end{aligned}$$

Considering the remaining figure as a (4×6) we have

$$u_2 = .02857P_1 + .17143(P_2 + u_3) = 3.950.$$

Lastly

$$u_1 = .16667(P_1 + u_2) = 3.462.$$

(b) Alternate procedure for u_1 and u_2 . After finding u_4 and u_3 , we take coefficients from Table IV and write

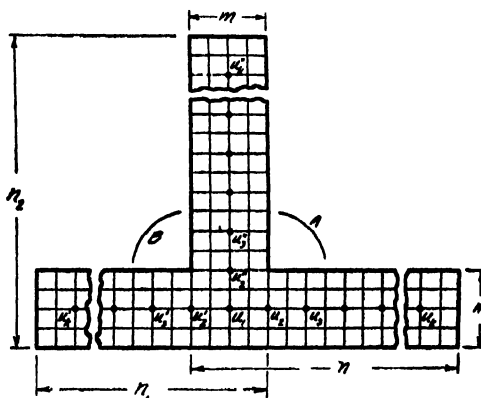


FIG. 11.

$$\begin{aligned} u_2 &= .02941P_1 + .17647P_2 \\ &\quad + .02941(P_3 + u_3) = 3.939, \\ u_1 &= .17155P_1 + .02941P_2 \\ &\quad + .00490(P_3 + u_3) = 3.460. \end{aligned}$$

The exact values of u_4 , u_3 , u_2 , and u_1 are 4.7536, 4.3820, 3.9512, 3.4657, which differ but slightly from those given by our method. From the key values all remaining values can be calculated by the Liebmann formula.

17. COMPOSITE RECTANGULAR AREAS

(a) Angles. The proposed method and tables are applicable to composite rectangular areas. Consider the angle of Fig. 10. This configuration may be viewed as the composite of two rectangles ($m \times n$) and ($m \times n_1$). The key values

$u_1, u_2 \dots u_k$ in rectangle ($m \times n$) could be found by means of the coefficients in the tables if u_2' were known. Similarly $u_1, u_2' \dots u_k'$ in the rectangle ($m \times n_1$) could be calculated if u_2 were known. The necessary key values u_2 and u_2' can be found from the following two equations:

$$\left. \begin{aligned} u_2 &= C_1(P_1 + u_2') + C_2(P_2 + .5u_2') \\ &\quad + C_3P_3 + \dots + C_kP_k, \quad (a) \\ u_2' &= C_1'(P_1' + u_2) + C_2'(P_2' + .5u_2) \\ &\quad + C_3'P_3' + \dots + C_k'P_k', \quad (b) \end{aligned} \right\} (17.1)$$

in which the coefficients C and C' are taken directly from the tables, and most perimeters are calculated from the basic expression

$$P = .5 \sum \text{known } a\text{'s} + \sum \text{known } c\text{'s}, \quad (17.2)$$

using only the boundary values on the angle.

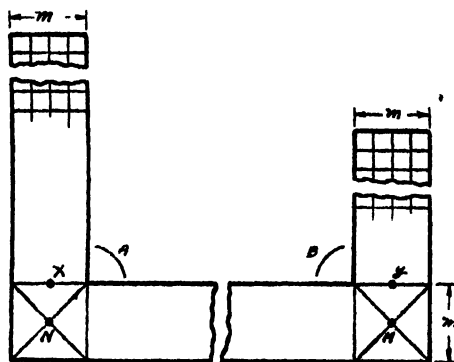


FIG. 12.

We denote the results from (17.1) by

$$\left. \begin{aligned} u_2 &= f(C, C', P, P'), \quad (a) \\ u_2' &= f_1(C, C', P, P'). \quad (b) \end{aligned} \right\} (17.3)$$

(b) T-figures. A configuration such as shown in Fig. 11 may be viewed as the composite of two overlapping angles. Observing that P_1 in angle A equals P_1 in the T-figure plus u_2' , and that P_1' in angle B equals P_1' in the T-figure plus u_2 we have by (17.3)

$$\left. \begin{aligned} u_2 &= f(C, P, C'', P'', u_2'), \\ u_2' &= f_1(C', P', C'', P'', u_2). \end{aligned} \right\} (17.4)$$

Upon solution the problem is reduced to angles.

An alternate procedure is to view Fig. 11 as consisting of three overlapping rectangles of lengths n, n_1 , and n_2 . The resulting equations are:

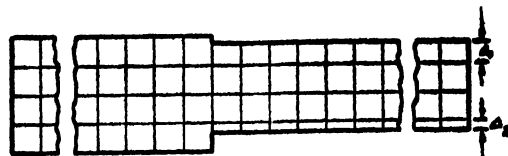


FIG. 13.

$$\left. \begin{aligned} u_2 &= C_1(P_1 + u_2' + u_2'') \\ &\quad + C_2(P_2 + .5u_2'') + C_3P_3 + \dots + C_kP_k, \quad (a) \\ u_2' &= C_1'(P_1' + u_2 + u_2'') \\ &\quad + C_2'(P_2' + .5u_2'') \\ &\quad + C_3'P_3' + \dots + C_k'P_k', \quad (b) \\ u_2'' &= C_1''(P_1'' + u_2 + u_2') \\ &\quad + C_2''(P_2'' + .5u_2 + .5u_2') \\ &\quad + C_3''P_3'' + \dots + C_k''P_k'', \quad (c) \end{aligned} \right\} (17.5)$$

Upon solution the figure is reduced to rectangles.

(c) Channels. A channel, Fig. 12, may be viewed as the composite of two overlapping angles A and B . We observe that

$$\left. \begin{aligned} (P_M)_{\angle A} &= (P_M)_{\text{channel}} + y, \quad (a) \\ (P_N)_{\angle B} &= (P_N)_{\text{channel}} + x. \quad (b) \end{aligned} \right\} (17.6)$$

Hence by (17.3)

$$\left. \begin{aligned} x &= f(C, C', P, P', y), \quad (a) \\ y &= f_1(C, C', P, P', x). \quad (b) \end{aligned} \right\} (17.7)$$

Upon solution we obtain x and y and the problem is reduced to angles. When the length $n > 3m$ the value y may be neglected in the calculation of x , and x may be neglected in the calculation of y . Equations (17.7) then reduce to (17.3). A figure having a Z-shape can be treated in the same manner.

(d) H or I shapes may be viewed as overlapping channels. The method is also applicable to rectangles of variable widths, Fig. 13. It should, however, be added that in complicated areas the method of block iteration may be shorter.⁴

SHORT BIBLIOGRAPHY

1. G. H. Shortly and R. Weller, "The numerical solution of Laplace's equation," J. App. Phys. 9, 334-348 (1938).
2. D. G. Christopherson and R. V. Southwell, "Relaxation methods applied to engineering problems," Proc. Roy. Soc. A168, 317-350 (1938). See also—H. W. Emmons, "The numerical solutions of heat conduction problems," Trans. A.S.M.E. 65, 607-615 (1943).
3. David Moskovitz, "The numerical solution of Laplace's and Poisson's equations," Quart. Appl. Math. 2, No. 2, 148-163 (July 1944).
4. Max M. Frocht, *Photoelasticity* (John Wiley and Sons, Inc., New York), Vol. II, Chap. 8, now in press.

Erratum

Correction of Diffraction Amplitudes for Lorentz and Polarization Factors

[J. App. Phys. 17, 285 (April 1946)]

M. J. BUERGER AND GILBERT E. KLEIN

THE values of the functions in Tables III and VIII are each misplaced in the region $90\text{--}179.9^\circ$ so that they are found at a value of the given argument 0.1° lower than the correct argument. The tables may be utilized in this region by subtracting 0.1° from the desired value of T and looking up the function for this altered (false) argument. For example, if one desires the function for 147.3° , it should be looked up under the altered (false) argument 147.2° .

The Thermoluminescence and Conductivity of Phosphors

ROBERT C. HERMAN AND CHARLES F. MEYER

Applied Physics Laboratory, Johns Hopkins University, Silver Spring, Maryland

(Received May 3, 1946)

A simple theoretical treatment is given for the emission of light and associated variations of dark current that occur on warming zinc sulphide and zinc silicate type phosphors previously irradiated with ultraviolet light at low temperatures. Reasonably good agreement is found between the calculated and observed glow curve for Willemite. The same type of treatment is applied to the case of infra-red illumination of phosphors at low temperatures.

1. INTRODUCTION

THE phenomena of thermoluminescence and the associated increase in conductivity in certain phosphors such as zinc sulphide and zinc silicate are sufficiently interesting to warrant further study. Recently there has arisen a great deal of interest in the storage and release of energy in solids, particularly with respect to infra-red-sensitive phosphors. Simple models based on the energy-band theory of solids have been shown to be in agreement with experimental data by various investigators.¹⁻³ It is now generally accepted that a phosphor is an insulating crystal which can be made luminescent by the addition

of a small concentration of impurity or activator atoms. Figure 1 shows the spectrum of energy states of electrons in the crystal. Between the uppermost filled band, *A*, and the next highest band, *B*, the conduction band which is normally empty, there are localized electron states, such as *C*, *D*, and *E*. The narrow energy states *C* and *D* which lie in the forbidden region are the impurity states associated with the activator atoms whereas the states *E* represent the so-called trapping states which are associated with lattice imperfections such as strains, cracks, surfaces, etc. The trapping states, *E*, are populated with electrons to an extent which depends on the previous history of the phosphor. Photoconductivity arises when an electron is raised into the conduction band from states *A*, *C*, or *E*

¹ F. Seitz and R. P. Johnson, J. App. Phys. 8, 84, 186, 246 (1937).

² N. F. Mott, Trans. Faraday Soc. 34, 822 (1938).

³ R. P. Johnson, J. Opt. Soc. Am. 29, 387 (1939).

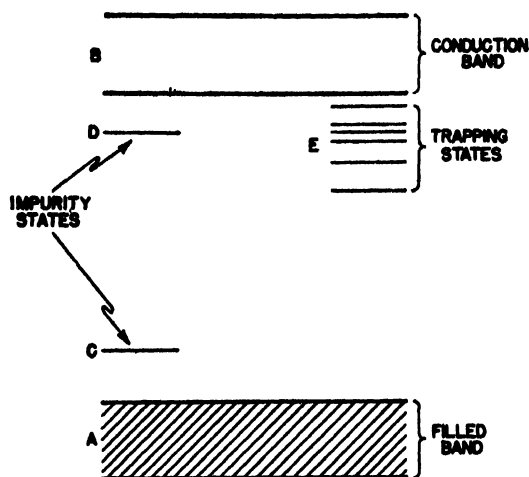


FIG. 1. Energy states of electron in phosphor.

by the absorption of a quantum of light. However, no photo-conductivity is observed when an electron is raised to an excited state at *D*.

In an excited phosphor, electrons return to the ground state by dropping from the conduction band to a state at *D* and then may make a radiative transition from *D* to *C*. Electrons in the conduction band may also be captured in the trapping states *E*. They can be released either thermally or by the absorption of radiation and thus return to the conduction band. The phenomena of phosphorescence and thermoluminescence are caused by the storage of electrons in excited states of the activator atoms and in the trapping states. Recently, Randall and Wilkins⁴ have discussed the phenomenon of thermoluminescence or glow and presented a series of significant experiments. The storage of light at low temperatures which can be recovered by warming is well known and has been discussed frequently. The glow curve for zinc silicate obtained by warming a phosphor or crystal previously illuminated at liquid air temperatures with ultraviolet light is accompanied by a large dark current peak at about the same temperature as the glow peak.⁵ This observation demonstrates that electrons thermally released from trapping states pass through the conduction band during the glow process. This con-

clusion is supported by the fact that a glow is produced in phosphorescent zinc silicate by the application of an electric field.⁶ As previously suggested, it would prove interesting to observe the variations of dark current in zinc-silicate type phosphors during the glow experiment.⁵ Also, dark current measurements would permit a study of trapping state distributions in non-fluorescent solids.⁴

It is the purpose of the present paper to discuss theoretically both the emission of light and variations in conductivity that occur in phosphors of the zinc silicate and zinc sulphide varieties. This has been done with a simplified model in which the only energy states considered are those in the uppermost filled band, the next highest normally empty conduction band and in the trapping state distribution. For the sake of simplicity, the analysis which follows deals with discrete sets of trapping states. This may be extended to various distributions as will be mentioned later.

2. LUMINESCENCE AND CONDUCTIVITY DURING WARMING

Let N_i be the number of trapping states per cm^3 at an energy ϵ_i below the conduction band, n_0 the number of positive holes per cm^3 in the uppermost filled band, n the number of electrons per cm^3 in the conduction band, and n_i the number of electrons per cm^3 in the trapping states in $d\epsilon_i$ at ϵ_i . If a phosphor of the type under discussion is irradiated with light of the proper wave-length at a sufficiently low temperature, the trapping states will become populated and remain so for considerable lengths of time so long as $\exp(-\epsilon_i/kT)$ is very small. With regard to the warming experiment, the initial condition of the crystal or phosphor chosen is that the number of electrons trapped in the states in $d\epsilon_i$ at an energy ϵ_i below the conduction band is $(n_i)_0$, $n=0$ and $(n_0)_0$ the number of holes, originally equal to the number of trapped electrons. In general, $n_0 = n + n_i$. If the crystal is allowed to warm up the light emission and variations in dark current may be found from the following set of equations:

⁴ J. T. Randall and M. H. F. Wilkins, Proc. Roy. Soc. A184, 366 (1945).

⁵ R. C. Herman and R. Hofstadter, Phys. Rev. 57, 936 (1940).

⁶ J. Frenkel, Tech. Phys. U.S.S.R. 5, 685 (1938).

$$dn_0/dt = -An_0n, \quad (1a)$$

$$\frac{dn}{dt} = -An_0n + \sum s_i \exp(-\epsilon_i/kT)n_i - \sum K_i[(n_i)_0 - n_i]n, \quad (1b)$$

$$\frac{dn_i}{dt} = -s_i \exp(-\epsilon_i/kT)n_i + K_i[(n_i)_0 - n_i]n, \quad (1c)$$

where A is the recombination coefficient and is equal to σV ; σ the cross section for capture by an impurity center of an electron in the conduction band and subsequent radiation; V the velocity of the conduction electron; s_i the frequency factor in sec^{-1} or the number of times per second an electron strikes the potential barrier in a trap; and K_i is similar to A but refers to the various sets of trapping states. De Groot⁷ has solved graphically a similar set of equations in his studies of the fluorescence and phosphorescence of zinc sulphide phosphors. Following Randall and Wilkins⁸ who point out that the effect of retrapping is not very large, we have solved the equations for the case $K_i = 0$ or:

$$dn_0/dt = -An_0n, \quad (2a)$$

$$\frac{dn}{dt} = -An_0n + \sum s_i \exp(-\epsilon_i/kT)n_i, \quad (2b)$$

$$\frac{dn_i}{dt} = -s_i \exp(-\epsilon_i/kT)n_i. \quad (2c)$$

Equation (2c), which expresses the rate at which trapped electrons leave the states in $d\epsilon_i$ at ϵ_i , can be integrated yielding:

$$n_i = (n_i)_0 \exp \left\{ -s_i \int_0^t \exp(-\epsilon_i/kT) dt \right\} \quad (3)$$

where $(n_i)_0$ is the number of electrons originally trapped in $d\epsilon_i$ at ϵ_i . If a constant rate of warming is used so that $T = \alpha t + T_0$, then $dT = \alpha dt$ and

$$n_i = (n_i)_0 \times \exp \left\{ -\frac{s_i}{\alpha} \int_{T_0}^T \exp(-\epsilon_i/kT) dT \right\}. \quad (4)$$

The general case involving electrons trapped in complicated trapping distributions can be

handled by a numerical integration of Eq. (4) over all trapping states.

In the case where there are sets of states at discrete energies the problem is much simpler and Eq. (4) becomes for a single trap energy:

$$n_i = N_i \exp \left\{ \frac{(s_i/\alpha k) \int_{x_0}^x x^{-2} e^{-x} dx \right\}, \quad (5)$$

where

$$x = \epsilon/kT, \quad x_0 = \epsilon/kT_0,$$

and it is assumed that N_i traps at the energy ϵ are filled with electrons when $T = T_0$. Equation (5) can be rewritten in the form

$$n_i = N_i \exp \{ -(s_i/\alpha k) \times [x^{-1} e^{-x} + Ei(-x) - f_0] \}, \quad (6)$$

where

$$-Ei(-x) = \int_x^\infty u^{-1} e^{-u} du$$

and

$$f_0 = x_0^{-1} \exp(-x_0) + Ei(-x_0).$$

Substitution of n_i from Eq. (6) into (2a) yields an equation that can be expressed in terms of known functions, but which requires numerical solution. It was found convenient to solve the differential equation numerically in the form:

$$\frac{d\eta}{dx} = -\frac{AN_1\epsilon}{\alpha k} \frac{\eta}{x^2} (\eta - \xi), \quad (7)$$

where

$$\eta = n_0/N_1, \quad \xi = n_i/N_1 \quad \text{and} \quad \xi = n/N_1 = \eta - \zeta.$$

The solution of these equations yields both the luminescent intensity, I , and the conduction current, i , in an applied electric field since $I \propto \eta \xi$ and $i \propto \xi$.

Randall and Wilkins⁸ have shown that for many of the phosphors the glow curves can be interpreted on the basis of a reasonably broad distribution of trapping states. For example, the broad glow curves for zinc sulphide phosphors originate in this manner. However, the three glow-curve peaks for Willemite+1 percent Mn seemed sufficiently sharp to be accounted for by a set of three discrete trap energies.

The light intensity and current have been computed for this case by solving Eq. (7) numerically using the value of the constants shown in Table I. The solutions to these equations are shown in Fig. 2 where η , ξ , and ζ are

⁷ W. de Groot, *Physica* 6, 275 (1939).

⁸ J. T. Randall and M. H. F. Wilkins, *Proc. Roy. Soc. A* 184, 390 (1945).

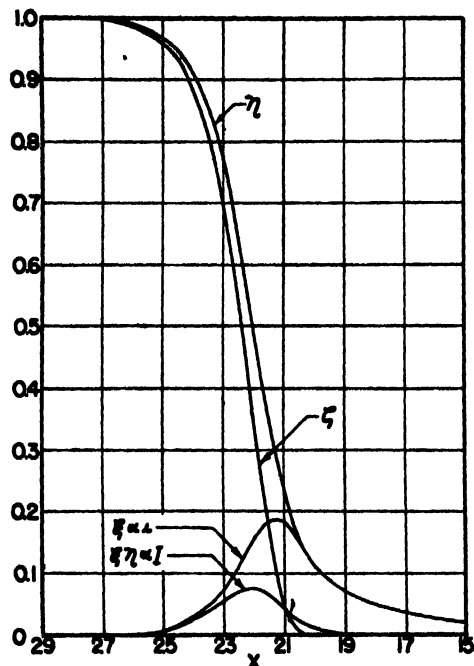


FIG. 2. Curves showing temperature dependence of luminescence and current during linear warming of a Willemite phosphor in an electric field, for the case of peak C (see Fig. 3).

plotted against x for the largest peak labeled C in Fig. 3. It is to be noted that the experimental fit seems to be good, particularly in that the long tail is predicted. For long times when the number of trapped electrons is small (or $\xi \rightarrow 0$) Eq. (7) reduces to a form having a solution of the type

$$\eta = \xi = [x/(1+cx)]. \quad (8)$$

Thus, for long times bimolecular decay applies. In the event that positive holes are originally present the decay would be more rapid and in fact exponential if there are a large number of holes present originally.

The conduction current is proportional to ξ and has not been plotted on a temperature scale since complete experimental data is not available for comparison. However, Herman and Hofstadter⁸ found a sudden increase in dark current at about -60°C which agrees with the predicted curve shown in Fig. 2.

Since the areas under the light emission curves are proportional to the number of trapped electrons, the relative number of trapping states at the three energies could be obtained. Since the peaks A and B overlap, it would be necessary

TABLE I.

Constants	Peak A	Peak B	Peak C
A (cm ² /sec.)	10^{-13}	10^{-13}	10^{-13}
s (sec. ⁻¹)	10^9	10^9	10^9
ϵ (ev)	0.223	0.291	0.500
N_1 (cm ⁻³)	1.4×10^{13}	$\sim 5 \times 10^{11}$	1×10^{13}
α (deg./sec.)*	5	5	3

* The experimental data of Randall and Wilkins (reference 4) shows a non-uniform rate of warming for this case.

to solve Eq. (7) in the form

$$\frac{d\eta}{dT} = -\frac{AN}{\alpha} \eta [\eta - \xi(T)],$$

where $N\xi(T)$ is the number of electrons trapped in both sets of levels. However, for the sake of simplicity each set of trapping states was handled separately and the result is shown in Fig. 3. Since overlapping light emission and current peaks are more difficult to analyze, it would be desirable to have a slow linear increase in temperature since this will lead to better resolution. This has been done by Williams and Eyring.⁹

3. LUMINESCENCE AND CONDUCTIVITY DURING INFRA-RED ILLUMINATION

Some time ago Gisolf¹⁰ performed a series of qualitative experiments with ZnS phosphors which showed that electrons in trapping states absorb infra-red radiation. The electrons can be freed by infra-red illumination and since changes in conductivity occur it is evident that the electrons pass through the conduction band. On first illuminating the phosphor the photo-current increases and then decreases as a consequence of recombination of electrons with positive holes. A more quantitative study by Reimann¹¹ shows that the resistance of a ZnS phosphor containing trapped electrons increases linearly with time when the phosphor is irradiated with infra-red light. The simple theory for sufficiently long time after beginning the illumination has been given by Reimann and others.^{12,13}

⁹ F. E. Williams and H. Eyring, "Mechanism of luminescence of solids," Bull. Am. Phys. Soc. 21, 17 (A) (Jan. 24, 1946). We wish to thank Professor H. Eyring and Mr. F. E. Williams for interesting discussions.

¹⁰ J. H. Gisolf, Ned. T. Natuurk 5, 289 (1938).

¹¹ A. L. Reimann, Nature 140, 501 (1937).

¹² N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, New York, 1940).

¹³ F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940).

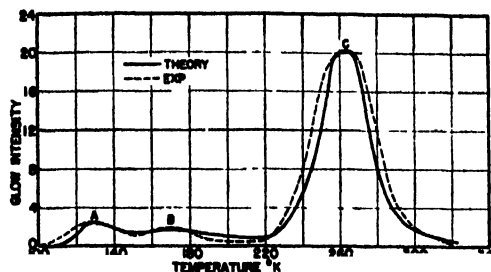


FIG. 3. Comparison of theoretical glow curve with experimental results of Randall and Wilkins (reference 4) for Willemite +1 percent Mn.

It is evident that the resistance of such a phosphor must decrease and then increase in a manner which depends upon the infra-red light intensity, the cross section of a trapped electron for an infra-red photon, the total number of trapped electrons, and the probability that a conduction electron will recombine with a positive hole. Since a general theory involving complicated trap distributions leads to tedious solutions, attention has been confined to the case of a single set of trapping states all at the same energy below the conduction band.

The equations are similar to those employed in the previous case except that the temperature is kept constant and the rate of release of electrons from trapping states is given by

$$\frac{dn_1}{dt} = -\frac{I_0}{h\nu}\sigma n_1, \quad (9)$$

where I_0 is the intensity of the infra-red light, σ the cross section of a trapped electron for an infra-red photon, and n_1 the number of trapped electrons. For constant intensity of illumination, I_0 , Eq. (9) yields

$$n_1 = N_1 \exp(-C_0 t), \quad (10)$$

where

$$C_0 = I_0 \sigma / h\nu,$$

and N_1 is the number of electrons originally trapped. Substitution of Eq. (10) into (2a) yields

$$\frac{dn_0}{dt} = -A n_0 [n_0 - N_1 \exp(-C_0 t)]. \quad (11)$$

The solution of Eq. (11) is

$$\frac{1}{\eta} = 1 - \frac{A N_1}{C_0} [Ei(-y) - Ei(-y_0)] e^y, \quad (12)$$

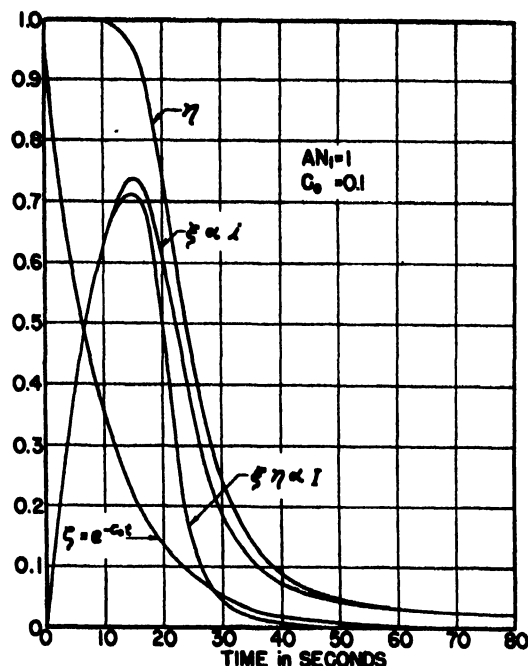


FIG. 4. Time dependence of luminescence and current for a phosphor in an electric field, that is illuminated with infra-red light at constant temperature.

where

$$y = (A N_1 / C_0) \exp(-C_0 t), \quad y_0 = A N_1 / C_0$$

and

$$\eta = n_0 / N_1.$$

When t is sufficiently large $\xi = (n/N_1)$ is equal to η and the resistance which is proportional to $(1/\xi)$ increases linearly with the time.

The solutions for the glow intensity and the current are shown in Fig. 4 for the case of $A N_1 = 1 \text{ sec.}^{-1}$ and $C_0 = 0.1 \text{ sec.}^{-1}$. The equations have also been solved for various values of $A N_1$ and C_0 and the resistance computed as a function of time. These results are shown in Fig. 5. It is apparent that if the infra-red intensity or the cross section is smaller, the resistance minimum is broader. On the other hand, if the total number of electrons originally trapped or the recombination coefficient is smaller, then the resistance begins to rise linearly at an earlier time. Since the resistance in the dark is finite, it is necessary to add a constant term to the conductivity calculated. However, if the peak current is large compared to the dark current, deviations from linearity will occur only at very long times.

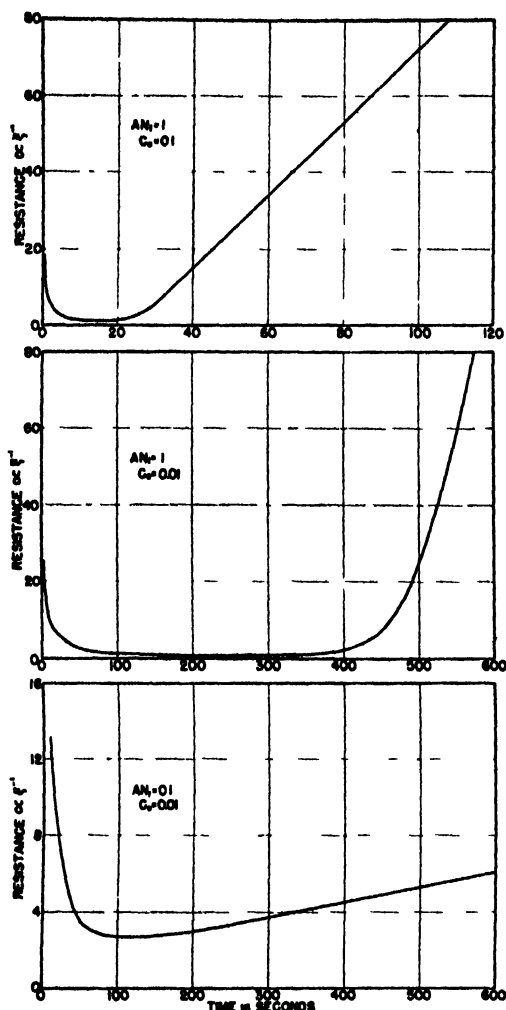


FIG. 5. Variation of resistance $\propto 1/\xi$ with time, for a phosphor in an electric field during infra-red illumination and the dependence on phosphor characteristics as well as light intensity.

4. APPENDIX

The exponential integral

$$-Ei(-x) = \int_x^\infty \frac{e^{-u}}{u} du$$

employed in the calculation of the glow intensity curves, etc., has been tabulated in various references¹⁴⁻¹⁶ for $0 \leq x \leq 15$ and $20 \leq x \leq 50$. We

¹⁴ *Tables of Sine, Cosine, and Exponential Integrals* (Federal Works Agency, WPA, 1940), Vol. 1 and 2.

¹⁵ *British Assoc. Adv. Sci. Math. Tables* (Cambridge University Press, New York, 1931), Vol. 1.

¹⁶ T. Akihira, *Sci. Pap. Inst. Phys. Chem. Res. Tokio*, 181 (June, 1929), Table No. 3.

TABLE II. Exponential integral $-Ei(-x) = \int_x^\infty (e^{-u}/u) du$.

x	$-Ei(-x)$	Δ_2
15.0	1.91863×10^{-8}	0.02178×10^{-8}
15.1	1.72520×10^{-8}	0.01956×10^{-8}
15.2	1.55133×10^{-8}	0.01756×10^{-8}
15.3	1.39502×10^{-8}	0.01581×10^{-8}
15.4	1.25452×10^{-8}	0.01419×10^{-8}
15.5	1.12821×10^{-8}	0.01276×10^{-8}
15.6	1.01466×10^{-8}	0.01146×10^{-8}
15.7	9.12573×10^{-9}	0.01030×10^{-8}
15.8	8.20784×10^{-9}	0.09013×10^{-9}
15.9	7.38008×10^{-9}	0.08818×10^{-9}
16.0	6.64050×10^{-9}	0.07229×10^{-9}
16.1	5.97321×10^{-9}	0.06727×10^{-9}
16.2	5.37319×10^{-9}	0.06042×10^{-9}
16.3	4.83359×10^{-9}	0.05434×10^{-9}
16.4	4.34833×10^{-9}	0.04886×10^{-9}
16.5	3.91193×10^{-9}	0.04390×10^{-9}
16.6	3.51943×10^{-9}	0.03949×10^{-9}
16.7	3.16642×10^{-9}	0.03550×10^{-9}
16.8	2.84891×10^{-9}	0.03191×10^{-9}
16.9	2.56331×10^{-9}	0.02872×10^{-9}
17.0	2.30643×10^{-9}	0.02580×10^{-9}
17.1	2.07535×10^{-9}	0.02321×10^{-9}
17.2	1.86748×10^{-9}	0.02088×10^{-9}
17.3	1.68049×10^{-9}	0.01876×10^{-9}
17.4	1.51226×10^{-9}	0.01688×10^{-9}
17.5	1.36091×10^{-9}	0.01519×10^{-9}
17.6	1.22475×10^{-9}	0.01366×10^{-9}
17.7	1.10225×10^{-9}	0.01227×10^{-9}
17.8	9.92022×10^{-10}	0.01106×10^{-9}
17.9	8.92850×10^{-10}	0.10031×10^{-10}
18.0	8.03609×10^{-10}	0.08843×10^{-10}
18.1	7.23311×10^{-10}	0.08041×10^{-10}
18.2	6.51054×10^{-10}	0.07233×10^{-10}
18.3	5.86030×10^{-10}	0.06509×10^{-10}
18.4	5.27515×10^{-10}	0.05859×10^{-10}
18.5	4.74859×10^{-10}	0.05263×10^{-10}
18.6	4.27466×10^{-10}	0.04741×10^{-10}
18.7	3.84814×10^{-10}	0.04267×10^{-10}
18.8	3.46429×10^{-10}	0.03835×10^{-10}
18.9	3.11879×10^{-10}	0.03454×10^{-10}
19.0	2.80783×10^{-10}	0.03042×10^{-10}
19.1	2.52729×10^{-10}	0.02924×10^{-10}
19.2	2.27599×10^{-10}	0.02453×10^{-10}
19.3	2.04922×10^{-10}	0.02263×10^{-10}
19.4	1.84508×10^{-10}	0.02037×10^{-10}
19.5	1.66131×10^{-10}	0.01835×10^{-10}
19.6	1.49589×10^{-10}	0.01651×10^{-10}
19.7	1.34698×10^{-10}	0.01484×10^{-10}
19.8	1.21291×10^{-10}	0.01337×10^{-10}
19.9	1.09221×10^{-10}	0.01204×10^{-10}
20.0	9.83555×10^{-11}	0.01083×10^{-10}

have been unable to find tabulated values of $-Ei(-x)$ in the range $15 < x < 20$. These have been calculated using the asymptotic expansion

$$-Ei(-x) = e^{-x} \left[\frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \frac{3!}{x^4} + \frac{4!}{x^5} - \dots \right]$$

and are given in Table II together with second differences, Δ_2 , in the hope that they may be found useful by others.

Oscillation Conditions in Single Tuned Amplifiers

WILLIAM R. FAUST AND HUGO M. BECK
Naval Research Laboratory, Washington, D. C.
 (Received May 31, 1946)

An application of the calculus of finite differences has been made in obtaining the complete expression for the voltage gain of an n -stage amplifier having identical grid-to-plate impedances and plate loads when driven by a generator of any given internal impedance. An application of these results was made in determining the conditions of oscillation in a multi-stage amplifier using single-tuned circuits. It was found that there is a certain minimum grid-to-plate capacitance required to cause oscillation. If the actual grid-to-plate capacitances are below this value, the system will not oscillate. There also exists a region of stable gain, zero to $2^{1/n}$, wherein it is impossible to make the amplifier oscillate even with grid-to-plate capacitance arbitrarily large.

1. INTRODUCTION

THE calculus of finite differences has been applied to cascaded systems of vacuum tube amplifiers,¹ but apparently the full power and usefulness of the method has not been realized. It is obvious that a system of cascaded amplifiers lends itself quite readily to the methods of finite differences. The equivalent circuit of multitube amplifiers resembles, to a very great extent, an artificial transmission line made up of Pi or T sections with the exception that the amplifier contains internal current generators. Nevertheless, the difference equations for the two systems are quite similar in form and become identical when the g_m (mutual conductance) of the tubes vanishes.

The purpose of this paper is twofold: First, that of obtaining a complete expression for the voltage gain of a multi-stage amplifier. Such an expression describes the effect of the feedback between the grid and plates of the various stages. Second, that of applying the general expression for voltage gain to specific cases and using it to compute the conditions necessary for sustained oscillations to exist in a multi-stage, single-tuned amplifier.

2. DEVELOPMENT OF FUNDAMENTAL RELATIONS

The circuit of Fig. 1 is a cascaded multi-stage amplifier driven by a generator of e.m.f. e_{in} and internal impedance Z_0 with identical plate load impedances Z_1 . The circuit of Fig. 2 is an

equivalent circuit of Fig. 1, wherein Z_1 represents the grid to plate impedance which is assumed to be perfectly general at the present time. For simplicity the circuit of Fig. 2 will be shown as in Fig. 2a, where Z_2 is the parallel combination of R_p and Z_2' , i.e.,

$$Z_2 = R_p Z_2' / (Z_2' + R_p).$$

The voltage relationship between the k and $k+1$ stages is

$$e_{k+1} - Z_1 i_k - e_k = 0, \quad (1)$$

and the currents at the plate of the k th stage are related by

$$i_k - i_{k-1} - e_k / Z_2 - g_m e_{k-1} = 0. \quad (2)$$

Equations (1) and (2) are the difference equations of the amplifier for which a solution satisfying the boundary conditions must be found. To find a solution, it is assumed that

$$e_k = U(\exp \gamma_1 k) + V(\exp -\gamma_2 k), \quad (3)$$

$$i_k = W(\exp \gamma_1 k) + X(\exp -\gamma_2 k), \quad (4)$$

where U , V , W , and X are complex constants and γ_1 and γ_2 are the complex propagation constants of the system. There are not four independent

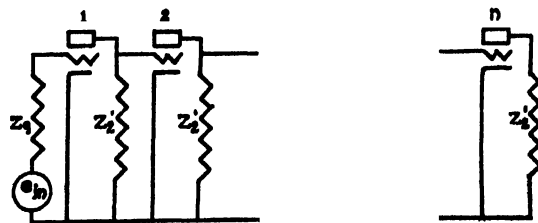


FIG. 1. Radiofrequency circuit of n -stage amplifier.

¹ A. C. Bartlett, *The Theory of Electrical Artificial Line and Filters* (John Wiley and Sons, Inc., New York, 1931), pp. 143-150.

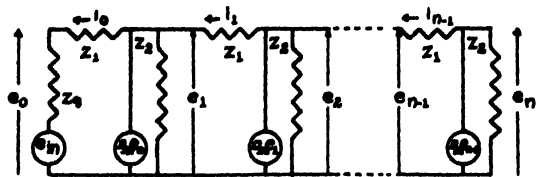
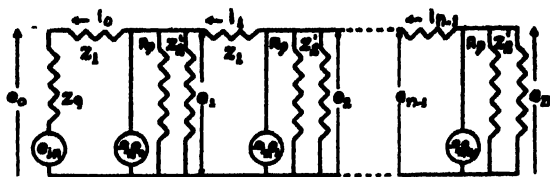


FIG. 2. Equivalent circuits. (a) Equivalent circuit of Fig. 1. (b) Simplified equivalent where $Z_s = R_p Z_2' / (Z_2' + R_p)$.

constants U, V, W, X but only two, as a relationship between U and W , and V and X can be found. By substituting the assumed solutions (3) and (4) into Eq. (1), it follows that

$$(\exp \gamma_1 k)[U(-1 + \exp \gamma_1) - Z_1 W] + (\exp -\gamma_2 k)[V(-1 + \exp -\gamma_2) - Z_1 X] = 0. \quad (5)$$

Since this expression must hold for all values of k , it is necessary that the coefficients of $\exp \gamma_1 k$ and $\exp \gamma_2 k$ vanish independently, i.e.,

$$W = U[(\exp \gamma_1) - 1]/Z_1, \quad (6)$$

$$X = V[(\exp -\gamma_2) - 1]/Z_1. \quad (7)$$

Equation (4) then becomes

$$i_k = U(-1 + \exp \gamma_1) \times (\exp \gamma_1 k)/Z_1 + V(-1 + \exp -\gamma_2) \times (\exp -\gamma_2 k)/Z_1. \quad (8)$$

To determine the "propagation constants" γ_1 and γ_2 so that the assumed solutions will satisfy the difference equations, substitute the solutions (3) and (8) into Eq. (2), which then becomes

$$U\{[-1 + \exp \gamma_1][\exp \gamma_1 k]/Z_1 - [-1 + \exp \gamma_1][\exp \gamma_1(k-1)]/Z_1 + [-\exp \gamma_1 k]/Z_2 - g_m[\exp \gamma_1(k-1)]\} + V\{[-1 + \exp -\gamma_2][\exp -\gamma_2 k]/Z_1 + [-1 + \exp -\gamma_2][\exp -\gamma_2(k-1)]/Z_1 - [\exp -\gamma_2 k]/Z_2 - g_m[\exp -\gamma_2(k-1)]\} = 0. \quad (9)$$

Because this must be satisfied for every value of k , it is necessary that the coefficients of U and V vanish; and since U and V are assumed not to equal zero, it follows that

$$\exp \gamma_1 = 1 + Z_1/2Z_2 \pm [Z_1/Z_2 + (Z_1/2Z_2)^2 + g_m Z_1]^{\frac{1}{2}}, \quad (10)$$

and

$$\exp \gamma_2 = 1 + Z_1/2Z_2 \pm [Z_1/Z_2 + (Z_1/2Z_2)^2 + g_m Z_1]^{\frac{1}{2}}/(1 - g_m Z_1). \quad (11)$$

The exact choice of the signs in (10) and (11) will be determined later. It should be pointed out, however, that $(\exp \gamma_1) \neq (\exp -\gamma_2)$ as such a choice of signs would lead to a solution not consistent with the physical facts. It has been shown that Eqs. (3) and (8) are solutions of the difference equations where γ_1 and γ_2 are as defined above and U and V determined by the boundary conditions.

3. BOUNDARY CONDITIONS

The boundary conditions are very similar to those for a transmission line and should require but little discussion. It will be assumed that the amplifier has n stages, so the corresponding boundary condition is simply

$$i_n = 0, \quad (12)$$

from which it follows that

$$V = -U[\exp n(\gamma_1 + \gamma_2)] \times [-1 + \exp \gamma_1]/(-1 + \exp -\gamma_2). \quad (13)$$

Since the amplifier is driven by a generator of e.m.f. e_{in} and internal impedance Z_0 , the corresponding boundary condition is

$$e_0 = e_{in} + Z_0 i_0. \quad (14)$$

By the use of Eqs. (1), (8), and (13) Eq. (14) becomes

$$e_{in} = U\{1 - [\exp n(\gamma_1 + \gamma_2)] \times [-1 + \exp \gamma_1]/(-1 + \exp -\gamma_2) + -Z_0(-1 + \exp \gamma_1) \times [1 - \exp n(\gamma_1 + \gamma_2)]/Z_1\}. \quad (15)$$

Consequently, the expression for the voltage gain between the generator and the k th stage becomes

$$\frac{e_k}{e_{in}} = \frac{[\exp \gamma_1 k] - [\exp n(\gamma_1 + \gamma_2)][-1 + \exp \gamma_1][\exp -\gamma_2 k]/[-1 + \exp -\gamma_2]}{1 - \{[\exp n(\gamma_1 + \gamma_2)][-1 + \exp \gamma_1]/[-1 + \exp -\gamma_2]\} - \left\{ \frac{Z_0}{Z_1} [-1 + \exp \gamma_1][1 - \exp n(\gamma_1 + \gamma_2)] \right\}} \quad (16)$$

Usually the over-all gain of the system is of interest rather than the gain up to a certain stage, so the expression for e_n/e_{in} will be considered. This is obtained from (16) simply by setting $k=n$,

$$\frac{e_n}{e_{in}} = \frac{(\exp -\gamma_2) - (\exp +\gamma_1)}{[\exp -n\gamma_1][-1 + \exp -\gamma_2] - [\exp n\gamma_2][-1 + \exp \gamma_1]} - \frac{Z_0}{Z_1} [-1 + \exp \gamma_1][-1 + \exp -\gamma_2][(\exp -n\gamma_1) - (\exp n\gamma_2)] \quad (17)$$

4. SPECIAL CASES

Equation (17), as it stands, is of little practical value; and the exponentials must be replaced by their values in terms of the parameters of the system. For simplicity in the manipulation of the expression, it will be convenient to set

$$A = 1 + Z_1/2Z_2,$$

and

$$B = [Z_1/Z_2 + (Z_1/2Z_2)^2 + g_m Z_1]^{\frac{1}{2}},$$

so that

$$\exp \gamma_1 = A - B \quad \text{or} \quad \exp -\gamma_1 = (A + B)/(1 - Z_1 g_m),$$

and

$$\exp \gamma_2 = (A - B)/(1 - g_m Z_1).$$

The relationships may be simplified by setting $Z_0 = Z_2$. This is the usual case and should not detract from the utility of the results. Substitution of these values in (17) produces

TABLE I.

No. of stages, n		Gain
0	1	
1		$\frac{Z_2 - g_m Z_2 Z_1}{2Z_2 + Z_1 + g_m Z_2^2}$
2		$\frac{(1 - g_m Z_1)^2 Z_2^2}{(Z_1 + 3Z_2 + 2g_m Z_2^2)(Z_1 + Z_2)}$
3		$\frac{(1 - g_m Z_1)^3 Z_2^3}{Z_1^3 + 3Z_1^2 Z_2(2 + g_m Z_2) + Z_1 Z_2^2(10 + 8g_m Z_2 + g_m^2 Z_2^2) + Z_2^3(4 + 3g_m Z_2)}$
4		$\frac{(1 - g_m Z_1)^4 Z_2^4}{(Z_1^3 + 3Z_1^2 Z_2 + g_m Z_1 Z_2^2)(Z_1^3 + 5Z_1^2 Z_2 + 3g_m Z_1 Z_2^2 + 5Z_2^3 + 4g_m Z_2^4)}$
5		$\frac{(1 - g_m Z_1)^5 Z_2^5}{Z_1^3 + 5Z_1^2 Z_2(2 + g_m Z_2) + 2Z_1^3 Z_2^2(18 + 16g_m Z_2 + 3g_m^2 Z_2^2) + Z_1^2 Z_2^3(g_m^2 Z_2^2 + 56 + 18g_m^2 Z_2^2 + 63g_m Z_2) + 5Z_1 Z_2^4(7 + 8g_m Z_2 + 2g_m^2 Z_2^2) + Z_2^5(6 + 5g_m Z_2)}$
6		$\frac{(1 - g_m Z_1)^6 Z_2^6}{(Z_1^3 + 7Z_1^2 Z_2 + 4g_m Z_1 Z_2^2 + 14Z_1 Z_2^3 + 13g_m Z_1 Z_2^2 + 2g_m^2 Z_1 Z_2^3 + Z_2^4 + 6g_m Z_2^5) \times (Z_1^3 + 5Z_1^2 Z_2 + 2g_m Z_1 Z_2^2 + 6Z_1 Z_2^3 + 3g_m Z_1 Z_2^2 + Z_2^4)}$

$$\frac{e_n}{e_{in}} = \frac{2B(1-g_m Z_1)^n}{(A+B)^n(A+B-1) - (A-B)^n(A-B-1) - (A-B-1)(A+B-1)[(A+B)^n - (A-B)^n] \frac{Z_2}{Z_1}}, \quad (18)$$

which may be written in the form

$$\frac{e_n}{e_{in}} = \frac{2B(1-g_m Z_1)^n}{(A+B)^n[(A+B)+g_m Z_2] - (A-B)^n[(A-B)+g_m Z_2]}. \quad (19)$$

By expanding Eq. (19), it is easy to obtain

$$\frac{e_n}{e_{in}} = \frac{(1-g_m Z_1)^n}{(n+1) \left[A^n + \frac{n(n-1)A^{n-2}B^2}{3!} + \frac{n(n-1)(n-2)(n-3)A^{n-4}B^4}{5!} + \dots \right] + g_m Z_2 \left[nA^{n-1} + \frac{n(n-1)(n-2)A^{n-3}B^2}{3!} + \frac{n(n-1)(n-2)(n-3)(n-4)A^{n-5}B^4}{5!} + \dots \right]}. \quad (20)$$

Equation (20) is the general expression for the gain of a cascaded n -stage amplifier and, as such, is much too general to deduce any results. In Table I, Eq. (20) has been evaluated for various values of n ; the appropriate values of A and B as given previously have also been included.

That these equations are correct in the limit may be shown by letting Z_1 become very large; the general form of the expressions in Table I then becomes $(-g_m Z_2)^n$, which is the proper form for the gain of an n -stage amplifier.

5. CONDITIONS FOR OSCILLATION

It is now desired to determine what conditions must exist in order to produce sustained oscillation. Obviously the system under consideration is oscillating when a finite output $e_n \neq 0$ exists for $e_{in} = 0$. In order that the above conditions obtain, it is necessary that the denominators of the various expressions in Table I vanish. To apply the above results to a practical case, Z_1 will be taken as the plate to grid impedance of a vacuum tube which is mainly capacitive, i.e.,

$$Z_1 = 1/j\omega C. \quad (21)$$

The effective plate impedance Z_2 will be taken as that of a single tuned circuit consisting of an inductance L_0 , capacitance C_0 , and resistance, all in parallel; hence²

² S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), pp. 119-120.

$$Z_2 = R/(1+JQF), \quad (22)$$

where

$$Q = R(C_0/L_0)^{1/2},$$

$$F = (\omega/\omega_0 - \omega_0/\omega),$$

$$\frac{\omega_0}{2\pi} = \text{Resonant frequency of single-tuned circuit}$$

$$= 1/2\pi(L_0 C_0)^{1/2},$$

$$\frac{\omega}{2\pi} = \text{The frequency under consideration, and}$$

R = Dynamic resistance

$$= R'R_p/(R_p + R') \approx R' \text{ (for pentodes).}$$

One Stage

From Table I, it is seen that the conditions of oscillation are to be determined from

$$2Z_2 + Z_1 + g_m Z_2^2 = 0. \quad (23)$$

By substituting the values of Z_1 and Z_2 from Eqs. (21) and (22) and separating the real and imaginary parts, there is obtained

$$(QF)^2 + 2(R\omega C)(QF) - 1 = 0 \text{ (Real part),} \quad (24)$$

and

$$2(QF) + 2(R\omega C) - g_m R(R\omega C) = 0. \text{ (Imaginary part).} \quad (25)$$

A simultaneous solution of these two expressions results in

$$QF = \pm [(2 + g_m R) / (g_m R - 2)]^{1/2}, \quad (26)$$

and

$$C = \mp 2 / \omega R [(g_m R)^2 - 4]^{1/2}. \quad (27)$$

The negative sign must be chosen in Eq. (26) so that C will be positive in Eq. (27). The frequency of oscillation ω is determined from

$$QF = (\omega / \omega_0 - \omega_0 / \omega) Q = (\omega^2 - \omega_0^2) Q / \omega \omega_0 \\ = -[(2 + g_m R) / (g_m R - 2)]^{1/2}, \quad (28)$$

$$C = \frac{4Q}{R\omega_0 \{ [(g_m R + 2)^2 + 4Q^2 (g_m^2 R^2 - 4)]^{1/2} - (g_m R + 2) \}}. \quad (30)$$

It is interesting to note that if $g_m R \leq 2$, the system will never oscillate, as C becomes indefinitely large when $g_m R = 2$, and is imaginary if $g_m R < 2$. For a high stage gain, these results differ but little from previous results.^{3, 4} For an amplifier of low gain ($g_m R$ of the order of 2) and low Q , there is a wide difference; and these results indicate that the system will be intrinsically stable. This situation is illustrated in the curves of Fig. 3 in which Eqs. (24) and (25) are plotted with (QF) as the independent variable and $(g_m R)$ a parameter. Equation (24) is a hyperbola and (25) is a straight line whose slope is dependent upon the parameter $g_m R$. The intersection of these two curves represents the simultaneous solution of the real and imaginary parts of the defining equation. In order to obtain positive values of $(R\omega C)$, the intersection of the two curves must occur in the second quadrant. Since the asymptote to the hyperbola corresponds to the straight line with $g_m R = 2$, it follows that there can be no real simultaneous solution of the two equations between $g_m R = 0$ and $g_m R = 2$. Consequently, the system will be inherently stable within these limits.

Two Stages

The defining equation in this case is

$$(Z_1 + Z_2)(Z_1 + 3Z_2 + 2g_m Z_1^2) = 0. \quad (31)$$

³ B. J. Thompson, "Oscillation in tuned radio frequency amplifiers," Proc. I.R.E., 19, 421-437 (1931).

⁴ J. R. Nelson, "Discussion of oscillation in tuned radio frequency amplifiers," Proc. I.R.E. 19, 1281 (1931).

which gives

$$\omega = \frac{\omega_0}{2Q} \left\{ - \left[\frac{(g_m R + 2)}{(g_m R - 2)} \right]^{1/2} \pm \left[\frac{(g_m R + 2)}{(g_m R - 2)} + 4Q^2 \right]^{1/2} \right\}. \quad (29)$$

The positive sign of the second term in (29) must be chosen so as to produce a frequency of oscillation that has physical meaning. After ω is eliminated from Eqs. (27) and (29) there results,

The second factor may be separated into its real and imaginary parts giving

$$(QF)^2 + 3(R\omega C)(QF) - 1 = 0 \text{ (Real part)}, \quad (32)$$

and

$$2(QF) + 3(R\omega C) + 2g_m R(R\omega C) = 0. \text{ (Imaginary part)}, \quad (33)$$

which when solved simultaneously produces

$$QF = \pm [(2g_m R + 3) / (2g_m R - 3)]^{1/2}, \quad (34)$$

and

$$C = \frac{\mp 2}{R\omega [4(g_m R)^2 - 9]^{1/2}}. \quad (35)$$

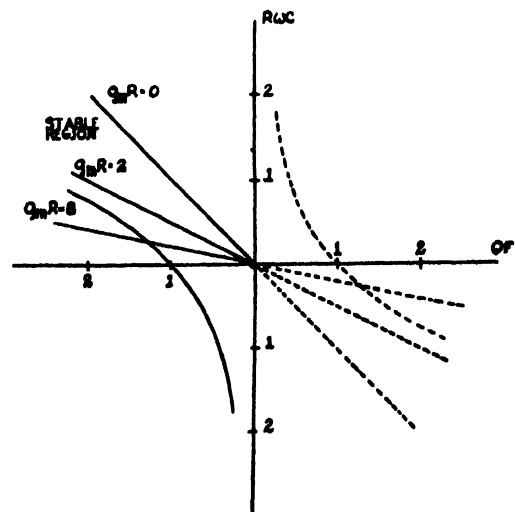


FIG. 3. Oscillation conditions for a single-tuned amplifier.

It is necessary again to choose the negative sign of (QF) so that a positive value of C will result. As before, ω may be determined from (34), thus:

$$\omega^2 - \omega_0^2 = -\frac{\omega_0 \omega \left[\frac{3 + 2g_m R}{2g_m R - 3} \right]^{\frac{1}{2}}}{Q \left[\frac{3 + 2g_m R}{2g_m R - 3} \right]^{\frac{1}{2}}},$$

or

$$C = \frac{4Q}{R\omega_0 \{ [(2g_m R + 3)^2 + 4Q^2(4g_m^2 R^2 - 9)]^{\frac{1}{2}} - (2g_m R + 3) \}}. \quad (37)$$

Again it is seen that there is a critical value of $g_m R$ below which it will be impossible to make the system oscillate. The critical value in this case is $g_m R = \frac{3}{4}$. For high stage gains and high Q 's, C approaches $1/\omega_0 g_m R^2$, which is the approximate value found by previous investigators. The inherently stable region is illustrated in the curves of Fig. 4. Here the asymptote to the hyperbola corresponds to the straight line with $g_m R = \frac{3}{4}$ so that the stable region lies between $g_m R = 0$ and $g_m R = \frac{3}{4}$.

The other part of the determining Eq. (31), $(Z_1 + Z_2) = 0$, produces no physically realizable conditions in this case and may, consequently, be neglected.

Three Stages

The definitive equation in this case is, from Table I,

$$Z_1^3 + 6Z_1^2 Z_2 + 3g_m Z_1^2 Z_2^2 + 10Z_1 Z_2^3 + 8g_m Z_1 Z_2^4 + g_m^2 Z_1 Z_2^4 + 4Z_2^3 + 3g_m Z_2^4 = 0, \quad (38)$$

which becomes, when separated into its real and imaginary parts,

$$(QF)^4 + 6R\omega C(QF)^3 + [10(R\omega C)^2 - 6](QF)^2 + 2R\omega C[2R\omega C - 3g_m R - 9](QF) + [1 - 10(R\omega C)^2 - 8g_m R(R\omega C)^2 - (g_m R)^2(R\omega C)^2] = 0 \text{ (Real)}, \quad (39)$$

and

$$-4(QF)^3 - 3R\omega C[6 + g_m R](QF)^2 + 4[1 - 5(R\omega C)^2 - 2g_m R(R\omega C)^2]QF + (R\omega C)\{6 - 4(R\omega C)^2 + 3g_m R[1 - (R\omega C)^2]\} = 0 \text{ (Imaginary)}. \quad (40)$$

Because an exact simultaneous solution of these equations is difficult to effect, an approximate

$$\omega = \frac{\omega_0}{2Q} \left\{ -\left[\frac{2g_m R + 3}{2g_m R - 3} \right]^{\frac{1}{2}} + \left[4Q^2 + \frac{2g_m R + 3}{2g_m R - 3} \right]^{\frac{1}{2}} \right\}, \quad (36)$$

wherein the positive sign of the radical has again been chosen to correspond with physical reality. Substitution of (36) in (35) then gives

solution is presented. If it is assumed that $(g_m R)$ is very much greater than one and $(g_m R)(R\omega C)$ is of the order of unity, then Eqs. (39) and (40) become

$$(g_m R)^2(R\omega C)^2 + 6g_m R(R\omega C)QF + [6(QF)^2 - (QF)^4 - 1] = 0, \quad (41)$$

and

$$(1 - Q^2 F^2)[3g_m R + 4QF] = 0. \quad (42)$$

It is easy to show that the only value of (QF) which leads to a real simultaneous solution of (41) and (42) is $(QF) = -1$. The resulting values of C are then

$$C = \frac{0.764}{g_m R^2 \omega} = \frac{1.526Q}{\omega_0 g_m R^2 [(1 + 4Q^2)^{\frac{1}{2}} - 1]}, \quad (43)$$

and

$$C = \frac{5.236}{g_m R^2 \omega} = \frac{10.472Q}{\omega_0 g_m R^2 [(1 + 4Q^2)^{\frac{1}{2}} - 1]}, \quad (44)$$

which are the results obtained by Thompson.

Because of the gross approximations involved in obtaining Eqs. (43) and (44), no region of intrinsic stability was found. By another method, it is possible to determine the stable region. If (QF) is eliminated between Eqs. (38) and (39), by some method such as Sylvester's, an equation of the twelfth degree in $(R\omega C)$ can be found. There is a theorem in algebra that states that the vanishing of the coefficient of the highest power term determines the infinite roots of the equation. From the previous discussion on one and two stages, it is apparent that this is just the condition that is desired. It can be shown that the coefficient of $(R\omega C)^{12}$ is $(3g_m R - 4)(g_m R - 4)(7g_m^2 R^2 + 16g_m R - 32)$, from which it follows that

$g_m R = 1.28$ is the only root having physical significance. The inherently stable region of stage gain ($g_m R$), therefore, lies between $g_m R = 0$ and $g_m R = 1.28$.

Four Stages

Table I indicates that the conditions of oscillation are to be determined from

$$(Z_1^2 + 3Z_1Z_2 + g_m Z_1Z_2^2 + Z_2^3)(Z_1^2 + 5Z_1Z_2 + 3g_m Z_1Z_2^2 + 5Z_2^3 + 4g_m Z_2^3) = 0. \quad (45)$$

The first quadratic on the left-hand side of (45), when separated into its real and imaginary parts, becomes

$$(QF)^2 + 3(R\omega C)(QF) + (R\omega C)^2 - 1 = 0 \text{ (Real part),} \quad (46)$$

and

$$2(QF) + 3(R\omega C) + (g_m R)(R\omega C) = 0 \text{ (Imaginary part).} \quad (47)$$

A simultaneous solution of these two equations yields

$$QF = \pm \frac{(3 - 2g_m R)}{[(g_m R)^2 - 5]^{\frac{1}{2}}}, \quad (48)$$

and

$$C = \frac{\mp 2}{R\omega[(g_m R)^2 - 5]^{\frac{1}{2}}}. \quad (49)$$

The angular velocity is found from Eq. (48) to be

$$\omega = \frac{\omega_0}{2Q} \left\{ \pm \frac{(g_m R - 3)}{[(g_m R)^2 - 5]^{\frac{1}{2}}} + \left[\frac{(g_m R - 3)^2}{(g_m R)^2 - 5} + 4Q^2 \right]^{\frac{1}{2}} \right\}, \quad (50)$$

where again the positive sign of the radical resulting from the solution of the quadratic for ω must be chosen to correspond with physical reality. To produce a physically realizable value of C , it is necessary to choose the positive sign, which in turn requires that QF be negative; consequently the negative sign of the first quantity within the brackets of (50) must be used. Elimination of ω from (49) results in

$$4Q \omega_0 R \{ -(g_m R + 3) + [(g_m R - 3)^2 + 4Q^2((g_m R)^2 - 5)]^{\frac{1}{2}} \} \quad (51)$$

This value of C is then the only real value that produces oscillation, due to the vanishing of the first quadratic on the left of (45).

An exact solution of the second quadratic on the left-hand side of Eq. (45) is very difficult to obtain so only an approximate solution will be attempted. The real and imaginary parts of the quadratic in question are

$$3(QF)^2 + (R\omega C)(10 + 3g_m R)(QF) + (R\omega C)^2(5 + 4g_m R) - 1 = 0, \quad (52)$$

and

$$(QF)^2 + 5(R\omega C)(QF)^2 + [5R\omega C - 3](QF) - (R\omega C)(5 - 3g_m R) = 0. \quad (53)$$

By Sylvester's method, it is possible to eliminate (QF) from these two equations and obtain

$$(R\omega C)^4(5 + 4g_m R)(g_m^2 R^2 + 20g_m R - 25) + (R\omega C)^4(81g_m^4 R^4 - 432g_m^3 R^3 - 2565g_m^2 R^2 - 3140g_m R - 400) + (R\omega C)^2(108g_m^2 R^2 - 48g_m R - 320) - 64 = 0. \quad (54)$$

The first term of this expression is negligible, relative to the second since it is of the order of $(R\omega C)^6(g_m R)^3$, while the second is of order $(R\omega C)^4(g_m R)^4$. If the gain is high, then it is possible to neglect all the terms but those of highest degree in the coefficients of $(R\omega C)^4$ and $(R\omega C)^2$. Equation (54) then becomes

$$(R\omega C)^4 81g_m^4 R^4 + (R\omega C)^2 108g_m^2 R^2 - 64 = 0, \quad (55)$$

of which the only physically realizable solution is

$$C = 2/3g_m R^2 \omega. \quad (56)$$

It may be shown that corresponding to the above solution for C , $(QF) = -1$, so that

$$\omega = \omega_0 \{ -1 + [1 + 4Q^2]^{\frac{1}{2}} \}. \quad (57)$$

Elimination of ω between Eqs. (56) and (57) produces

$$C = \frac{2}{3\omega_0[(1 + 4Q^2)^{\frac{1}{2}} - 1]g_m R^2}. \quad (58)$$

Because of the rough approximations made in

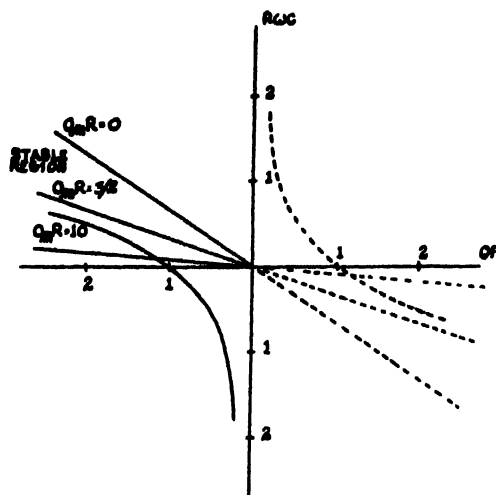


FIG. 4. Oscillation conditions for a two-stage single-tuned amplifier.

obtaining these results no stable region of gain is found; however, the infinite roots of $(R\omega C)$ are defined by the vanishing of the coefficient of the highest degree term in Eq. (54). The only physically realizable value of $g_m R$ is found to be $g_m R = 1.18$.

Five and Six Stages

Because of the complexity of these expressions, only partial solutions will be presented. Usually the least value of C that will cause oscillations is of the greatest interest so that only these results will be presented. By the use of approximations similar to those used for three stages it can be shown that for five stages

$$C = \frac{0.63}{g_m R^2 \omega} = \frac{1.26Q}{\omega_0 g_m R^2 [(4Q^2 + 1)^{1/2} - 1]}, \quad (59)$$

where (ω) is determined from the condition that $QF = -1$. The stable limits of gain lie between $g_m R = 0$ and $g_m R = 1.15$, approximately.

For six stages the corresponding value of C is

$$C = \frac{0.586}{\omega g_m R^2} = \frac{1.172Q}{g_m R [(4Q^2 + 1)^{1/2} - 1] \omega_0}, \quad (60)$$

in which ω is determined from $QF = -1$. The stable limits of gain are between $g_m R = 0$ and $g_m R = 1.12$, approximately.

6. CONCLUSIONS

The general expression for the voltage gain of an n -stage amplifier having identical but arbitrary plate impedances driven by a generator of any internal impedance has been obtained by the methods of the calculus of finite differences. The voltage gains, computed by this means reduce to the correct form as the grid-to-plate impedances of the tubes increase without limit.

An application of the general expression for the gain has been made in determining the criteria of oscillation in multi-stage, single-tuned amplifiers. These results check previous work for large values of Q , but also indicate a region of intrinsically stable stage gain wherein it is impossible to make the system oscillate even with a grid-to-plate capacitance arbitrarily large. The stable region of stage gain lies between zero and 2, 1.50, 1.28, 1.18, 1.15, and 1.12 for one, two, three, four, five, and six stages, respectively. An approximation, accurate within 10 percent of this sequence, is $(2)^{1/n}$ where n is the number of stages.

It is apparent from the discussion of oscillation conditions that if the actual inter-electrode capacity of the vacuum tubes in a cascaded amplifier is smaller than the least value calculated by the results derived above, the system can never oscillate because of feedback through the plate-to-grid capacitance and can be made to serve as a useful amplifier.

Calendar of Meetings

September

- 18-21 Illuminating Engineering Society, Quebec, Canada
- 19-21 American Physical Society, New York, New York
- 26-28 American Physical Society (Division of High Polymer Physics) and The Fiber Society (Charlottesville, Virginia) (Joint Meeting)
- 30-October 3 American Society of Mechanical Engineers, Boston, Massachusetts

October

- 3-5 National Electronics Conference, Inc., Chicago, Illinois
- 3-5 Optical Society of America, New York, New York
- 9-11 American Society of Civil Engineers, Kansas City, Missouri
- 25 American Mathematical Society, New York, New York

November

- 7-9 Conference on Electrical Insulation, Division of Engineering and Industrial Research, National Research Council, Baltimore, Maryland
- 14-16 Acoustical Society of America, Chicago, Illinois
- 29-30 American Physical Society, Minneapolis, Minnesota

December

- 2-6 American Society of Mechanical Engineers, New York, New York
- 5-7 Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction, Pittsburgh, Pennsylvania (Joint Meeting)
- 17 Institute of Aeronautical Sciences, New York, New York
- 26-28 Geological Society of America, Chicago, Illinois

Letter to the Editor

The Evaporation of Antimony

LOUIS HARRIS

Massachusetts Institute of Technology, Cambridge, Massachusetts*

June 19, 1946

THE aggregate formation observed in the evaporation of antimony has been found to have a counterpart in a simultaneous decrease in electrical resistivity of thin antimony deposits. The evaporation procedure has been modified as a result of these findings to produce films having lower electrical resistivity and higher thermal e.m.f. than obtained previously.

An unusual phenomenon was observed during studies on the evaporation of antimony. If the heating current was discontinued after the antimony was first observed to be condensing—as noted by an evaporation “meter” (see below)—the “meter” resistance continued to fall during a period of two to four minutes. For a number of runs (but not for every run) the decrease in resistance with time was found to follow the course of a first-order reaction. The phenomenon is not difficult to repeat, if the heating current is discontinued at a “meter” reading of about 5×10^8 ohms (estimated antimony thickness = 5×10^{-7} cm), the final reading of the “meter” will vary from 2000 to 10,000 ohms. (It will depend on the rate of heating before the heating current was discontinued.) The phenomenon has been observed for other types of meters than the one described here. It is best explained by assuming that the meter is recording a change to crystalline aggregates of antimony.¹

The meter was directly (15 cm) above the evaporation boat (molybdenum). Two cellulose nitrate films, supported on mica frames, were disposed to either side of the meter. By looking down into the bell jar at the back of the cellulose nitrate films, during the supposed aggregation process, one can observe the growing areas of aggregates, as has already been reported.^{2,3} (The growing areas may be observed also on the walls of the bell jar.)

We have modified the evaporating technique for antimony on the basis of these observations. We allow four minutes to elapse for the “crystallization” of the first deposit before proceeding further with the evaporation. The current through the boat is then increased gradually over a period of twenty minutes, until the boat is “red hot” and all the antimony has been evaporated. A vacuum of 10^{-4} mm Hg is maintained.

The progress of the evaporation was followed with the “meter” mentioned above.⁴ The meter consists of a glass strip 5 cm $\times$ 0.35 cm. Chromium-silver is evaporated[†] at the ends and No. 22 copper wire leads are “soft” soldered to the chrome-silver. Leads from terminals are brought out of the bell jar to a “volt-ohmist.” The electrical resistance of the meter serves as rough measure of the amount of metal deposit and of the resistivity of the metal deposit at the end of the evaporation process. Such a meter has been used for hundreds of evaporations and can be used for other metals besides antimony. The

antimony is removed from the meter after the evaporation, by solution with concentrated hydrochloric acid or by rubbing with a damp cloth.

The antimony deposits are a brilliant steel-grey and show a pattern of tiny crystals to the unaided eye. The resistivity of the antimony deposits on cellulose nitrate films varies from 119×10^{-6} to 60×10^{-6} ohm-cm, for antimony thicknesses from 0.48 to 2.2×10^{-4} cm. The thicknesses were determined from weight measurements and assuming macroscopic density. The thermal e.m.f. versus copper of the antimony deposits varied from 33 to 38.6 microvolt per degree C.

If the initial rate of antimony deposition is too high, or if sufficient time is not allowed to elapse, then the final deposit does not show the crystalline pattern, though it will still have the brilliant steel-grey appearance. Such deposits have a more uniform appearance; however, the electrical resistivity of such films is usually greater than 120×10^{-6} ohm-cm and the thermal e.m.f. versus copper may fall to less than 30 microvolt per degree C.

* The work described here was carried out in whole or in part under contract OEMr-1147 between the Office of Scientific Research and Development and the Massachusetts Institute of Technology.

¹ L. Marton and L. I. Schiff, J. App. Phys. 12, 759 (1941).

² Hans Murmann, Zeits. f. Physik 54, 755 (1929).

³ R. Bowling Barnes and M. Cerny, Phys. Rev. 38, 338 (1931).

⁴ The meter used here is a modification of one used originally in this laboratory by A. Stockfleth (A. Stockfleth, Massachusetts Institute of Technology Ph.D. Thesis, November, 1942).

[†] I am indebted to Professor John D. Strong for evaporating the silver-chromium terminals for the meter.

Here and There

New “Microwave Tower”

Federal Telephone and Radio Corporation of Newark, New Jersey, is building a 300-foot “microwave tower” for Federal Telecommunication Laboratories at Nutley, New Jersey. It is believed that such towers will be headquarters in cities and towns for simultaneous use of many microwave radio functions, including f-m broadcasting, Pulse Time Modulation broadcasting, television, police radio networks, communication with mobile units, radar applications, aerial navigation, and the interception of illegal transmissions. At the ground-breaking ceremony Colonel Sothenes Behn, President of International Telephone and Telegraph Corporation, stated, “Various broadcasting services in important communities will find it much more efficient to cooperate in building one great tower to be shared by all functions utilizing microwaves than for each service to build its own tower. Joint financing would make it possible to build a tower superior to any that could be afforded by individual services.”

Federal’s microwave tower will be an experimental laboratory, first of its kind ever to be built. Microwave experiments heretofore have been conducted in tall office buildings and other structures not well adapted to the purpose.

Ground has been broken also for the newest addition to Federal's existing Nutley laboratories, which will increase the present floor space of the laboratories by 65,000 square feet.

Meteorology Experts Available

The American Meteorological Society, through its president, Professor H. G. Houghton of Massachusetts Institute of Technology, wishes to call to the attention of American colleges the fact that a large number of carefully selected college students were given training in professional meteorology at a few selected universities in this country, preparatory to their employment by the Army and Navy Meteorological Services. A number of these men, before entering the meteorological field, had done a considerable amount of graduate work in physics and are now available for teaching and research in both these fields.

In most of our colleges which offer non-professional courses in meteorology to meet the needs of students in geography, geology, and agricultural subjects, it has been standard practice to have these non-professional courses taught by members of various departments. The return of many of the professionally trained service meteorologists offers an opportunity for our colleges to obtain the services of instructors who are capable of doing professional work in a physics department and at the same time are qualified as experienced and trained professional meteorologists to present stimulating non-professional courses in meteorology on a somewhat higher plane than may have been possible in the past.

The Society is anxious to cooperate with American colleges in providing information concerning these returning weather officers that might be helpful in placing them in suitable teaching positions. To that end, files have been kept which contain detailed information on the education and practical experience of these men. Inquiries may be directed to the Executive Secretary, American Meteorological Society, Five Joy Street, Boston 8, Massachusetts.

Books for Europe

The American Book Center for War Devastated Libraries, Inc., is collecting and shipping abroad scholarly books and periodicals which will be useful in research and necessary in the physical, economic, social, and industrial rehabilitation and reconstruction of Europe and the Far East. The Center cannot purchase books and periodicals; it must depend upon gifts from individuals, institutions, and organizations.

What is needed: Shipping facilities are precious and demand that all materials be carefully selected. Emphasis is placed upon publications issued during the past decade, scholarly books which are important contributions to their fields, periodicals (even incomplete volumes) of significance, fiction and non-fiction of distinction. All subjects—history, the social sciences, music, fine arts, literature, and especially the sciences and technologies—are wanted.

What is not needed: Textbooks, out-dated monographs, recreational reading, books for children and young people, materials of purely local interest, popular

magazines, popular non-fiction of little enduring significance, etc.

How to ship: All shipments should be sent prepaid via the cheapest means of transportation to The American Book Center, care of The Library of Congress, Washington 25, D. C. The Center cannot accept material which is sent collect. When possible, periodicals should be tied together by volume. It will be helpful if missing issues are noted on incomplete volumes.

New Section at Midwest Research Institute

The Midwest Research Institute, Kansas City, Missouri, recently announced the formation of its Engineering Mechanics Section. Previously, work in this field was done under the supervision of the Applied Physics Section. The new section is equipped with a large physical testing laboratory, machine shop, and electromechanical laboratory. Problems within the scope of the group include structural and mechanical design, vibrations, fluid flow, electron application in engineering, and servo-mechanisms.

Martin Goland, formerly of the Structures Department of the Curtiss-Wright Research Laboratory, Buffalo, New York, has been appointed chairman of the new section. Clayton O. Dohrenwend, formerly chairman of the Mechanics Departments at the Illinois Institute of Technology and the Armour Research Foundation, will serve as research consultant and senior scientist.

Food Packages for Europe

The Cooperative for American Remittances to Europe, Inc. (known as CARE) has announced a new service for sending food packages to Europe. CARE is a non-profit corporation organized by 24 relief agencies, with the support and approval of the President's War Relief Control Board.

CARE accepts orders for individuals, groups, or organizations in Austria, Czechoslovakia, Finland, France, Italy, Netherlands, Norway, and Poland, and recommends the purchase of a 10-in-1 package of Army rations for \$15 each, with delivery overseas guaranteed. Orders when received here are rushed to Europe and filled overseas from a stockpile of CARE food packages already over there. When the food is delivered, CARE's overseas representative will get a signed receipt which will be sent to the donor. If a package cannot be delivered within a reasonable length of time, CARE will refund the full purchase price to the donor. A typical CARE package contains such products as Borden's butter, Nescafe, Swift's canned ham, Del Monte corn, Welch's jam, Hormel's sliced bacon, etc.

CARE's order blanks are available at banks and are also distributed by relief, fraternal, professional, and social organizations, but if you have difficulty in getting one, write to CARE, 50 Broad Street, New York 4, New York.

Annual Convention of I.E.S.

The Illuminating Engineering Society goes to Quebec September 18-20 for its Victory Convention, sharing with its Canadian membership this annual event which during the war years was limited to technical conferences.

Journal of Applied Physics

Volume 17, Number 10

October, 1946

Supplementary Bibliography of Electron Microscopy

BY MARY E. RATHBUN, MARY J. EASTWOOD, AND ORLAN M. ARNOLD

Engineering Division, Chrysler Corporation, Detroit, Michigan

IN an extensive survey of the literature on the subject of electron microscopy a number of publications have been discovered which were not included in the excellent bibliographies of C. Marton and S. Sass.¹ In view of the request expressed by a number of people interested in electron microscopy to present a list of these articles which have technical points of interest bearing on technique, preparation of samples, instrumentation, and results, the authors have prepared this supplementary bibliography in two alphabetized sections. The first group includes publications which have information on various phases of electron microscopy. The second section gives the patent literature that has appeared to date. The presentation of the latter section has been prompted by the expressed opinion that many working in this field of research would like to become familiar with different stages of development of the electron microscope as reflected in both foreign and domestic patents. The domestic patents have been checked with The Official Gazette of the United States Patent Office and the foreign patents with the information in the Chemical Abstracts. For the convenience of the reader, translations have been made for the titles of publications and patents appearing in foreign languages.

The authors wish to acknowledge the aid of Eleanor V. Wright, Librarian,* in helping to locate a number of the publications reviewed for this bibliography.

ELECTRON MICROSCOPE REFERENCES NOT INCLUDED IN BIBLIOGRAPHY OF C. MARTON AND S. SASS

- Ahearn, A. J. and Becker, J. A. Electron microscope studies of thoriated tungsten. *Phys. Rev.* **54**, 448-458 (1938).
- Barnes, R. B. and Burton, C. J. Electron microscope. *Am. Dyestuff Rep.* **31**, 254-262 (1942); correction **31**, 313-314 (1942).
- Barrett, C. S. Seeing metals at high magnification. *Mach. Design* **15**, 138-140 (November, 1943).
- Benjamin, M. Field emission microscope. *J. Soc. Glass Tech.* **24**, 93-96 (1940); abstracted in *Chem. Ind. Eng.* **59**, 314-315 (1940).
- Bensen, I. B. Electron microscopy. *Gen. Elec. Rev.* **47**, No. 12, 6-14 (1944).
- Boersch, H. The shadow-microscope, a new electron-supermicroscope. *Naturwiss.* **27**, 418 (1939).
- Brüche, E. Electron microscope. *Naturwiss.* **20**, 49 (1932).
- Brüche, E. Electron optics and electron microscopy. *AEG Mitt.* No. 2, 45-47 (February, 1934).
- Brüche, E. Ten years of development. *AEG-Forsch. Jahrb., Sonderheft, Übermikroskop* **7**, 2-8 (1940).
- Brüche, E. The two-pole system as aim of pure electrical image apparatus. *AEG-Forsch. Jahrb., Sonderheft, Übermikroskop* **7**, 9-14 (1940).
- Brüche, E. and Götz, E. Introduction of the object and the

¹ C. Marton and S. Sass, *J. App. Phys.* **14**, 522-31 (1943); **15**, 575-9 (1944); **16**, 373-8 (1945).

* Engineering Division, Chrysler Corporation, Detroit, Michigan.

- plate. *AEG-Forsch. Jahrb., Sonderheft, Übermikroskop* 7, 60-66 (1940).
- Brüche, E. and Johansson, H. Electron optics and electron microscope. *Naturwiss.* 20, 353-358 (1932).
- Burgers, W. G. Metallographic investigation with the electron microscope. *Polytech. Weekblad.* 33, 17-18, 38-40 (1939).
- Dautz, D. Supermicroscope and paint research. *Farbe u. Lack, Centralblatt* 402-403 (1938).
- Dorgelo, H. B. Introduction to electron microscopes of different types. *Nederlandsch Tijdschrift voor Natuurkunde* 7, 157-170 (1940).
- Gaede, The electron microscope. *Deut. Med. Wochschr.* 66, 858-860 (1940).
- Hendricks, S. B., Wildman, S. G., and McMurdie, H. F. Morphology of latex particles as shown by electron micrographs. *India Rubber World* 110, 297-300 (1944).
- Henneberg, W. The supermicroscope with electrostatic lenses. *Elektrotech. Zeits.* 61, 773-776 (1940).
- Hillier, J. The electron microscope in the determination of particle size characteristics. *Am. Soc. Test. Mat., Symposium on New Methods for Particle Size Det'n. in Subsieve Range 90-94* (1941).
- Hillier, J. Electron microscopy. *Can. Chem. Process Ind.* 28, 728-736 (1944).
- Hock, C. W. and McMurdie, H. F. Structure of the wool fiber as revealed by the electron microscope. *Am. Dyestuff Rep.* 32, 433-436, 451-454 (1944).
- Hoffmann, K. Characteristics of optical illustrations by means of electrons and applications (electron microscopy). *Kolloid Zeits.* 89, 59-76 (1939).
- Humbert, R. P. Particle shape and the behavior of clay as revealed by the electron microscope. *Bull. Am. Cer. Soc.* 21, 260-263 (1942).
- Jakus, M. A., Hall, C. E., and Schmitt, F. O. Electron microscope observations of clam muscle fibrils. *J. Am. Chem. Soc.* 66, 313-314 (1944).
- Jolley, E. E. Structure of the developed silver image as revealed by the electron microscope. *J. Phot. Soc. Am.* 8, 283-289 (1942).
- Kalden, H. Investigations with the supermicroscope. *Chem.-Zest.* 64, 129-133 (1940).
- Kinder, E. The magnetic sectional lens supermicroscope and some applications in colloid chemistry. *Kolloid Zeits.* 95, 326-36 (1941).
- Knaysl, G. Width and origin of bacterial flagella; electron microscope studies. *Science* 95, 406-407 (1942).
- Knoll, M. and Theile, R. Electron scanning for forming the image of the structure of surfaces and thin layers. *Zeits. f. Physik* 113, 260-280 (1939).
- Kronig, R. Theoretical basis of electron optics. *Nederlandsch Tijdschrift voor Natuurkunde* 7, 171-178 (1940).
- Loofbourou, J. R. Borderland problems in biology and physics; electron microscopy. *Rev. Mod. Phys.* 12, 331-333 (1940).
- Mahl, H. Metallurgical, metal chemical and bacteriological investigations with the electrostatic supermicroscope. *Chem. Fabrik* 14, 279-280 (1941).
- Mahl, H. Stereoscopic photographs with the electrostatic supermicroscope. *Naturwiss.* 28, 264 (1940).
- Malev, N. N. Electron microscope. *Uspekhi. Fis. Nauk.* 13, 367-384 (1933).
- Martin, L. C. The present position regarding the development of the electron microscope. *J. Soc. Glass Tech.* 24, 97-100 (1940).
- Morton, G. A. Survey of research accomplishments with the RCA electron microscope. *RCA Rev.* 6, 131-166 (1941).
- Morton, G. A. and Ramberg, E. G. Point projector electron microscope. *Phys. Rev.* 56, 705 (1939).
- Müller, E. W. Further observations with the field electron microscope. *Zeits. f. Physik* 108, 668-680 (1938).
- Pupko, S. L. Electron microscope. *Uspekhi. Fis. Nauk.* 24, 487-513 (1940).
- Recknagel, A. Defects of electron lenses. *Jahrb. AEG-Forsch., Sonderheft, Übermikroskop* 7, 15-22 (1940).
- Rees, A. L. G. Electron microscope. *J. Soc. Chem. Ind.* 60, 335-337 (1941).
- Rees, A. L. G. Electron microscopy. *Sci. J. Roy. College of Science* 12, 1-15 (1942).
- Rhea, H. E. The RCA electron microscope. *Wallerstein Labs. Commun.* 4, 99-106 (1941).
- Rhea, H. E. New electron microscope. *Science* 93, 357-358 (1941).
- Ross, S. Orientation of micelles in soap fibers. *J. Phys. Chem.* 46, 414-417 (1942).
- Schmitt, F. O. X-ray and electron microscope studies of the structure of collagen fibers. *Am. Leather Chem. Assn. J.* 39, 430-437 (1944); Discussion 437-441.
- Semmler, E. The structure of aluminum oxide films used in surface structure investigations by the contact film method (electron microscopy). *Zeits. f. Metallkunde* 34, 229-231 (1942)—Translation: Great Britain Air Ministry RTP, Translation No. 1880.
- Smith, P. C. and Picard, R. G. Electron microscopes. *Radio News* 32, 41-43 (1944) (November).
- Stüger, A. Electronic and scanning microscopes. *Elec. Rev. (London)* 124, 157-158 (1939).
- Stanley, W. M. Viruses and the electron microscope. *Chromica Botan.* 7, 291-294 (1943).
- Triau, L. The most recent realization of the electron microscope. *Rev. gén. élec.* 44, 841-847 (1938).
- Wilson, W. The electron microscope. *Electronic Eng.* 16, 414-420 (1944).
- Wilson, W. Electron microscope. Part I: The Instrument. *Elec. Rev. (London)* 134, 218-222 (1944). Part II: Specimen Mounting and Typical Results. *Elec. Rev. (London)* 134, 254-257 (1944).
- Zworykin, V. K. Electron microscopy in chemistry. *Electronics* 16, 64-68 (1943).
- Zworykin, V. K. Image formation by electrons. *Cold Spring Harbor Symposia Quant. Biol.* 9, 194-196 (1941); Discussion 196-197.
- Zworykin, V. K. Image formation by electrons. *Science in Progress* [3] 69-107 (1942).
- Zworykin, V. K., Morton, G. A., Ramberg, E. G., Hillier, J., and Vance, A. W. *Electron optics and the electron microscope* (John Wiley & Sons, Inc., New York, 1945), 766 pp.

Development and uses of the electron microscope. *Chemische Fabrik* 11, 478-479 (1938).
 Electron microscopes. *Electronics* 11, 30-33 (November, 1938).
 Electron microscopes for production, research and analysis. *Electronics* 17, 184 (June, 1944).
 Electron microscope: the Siemens electron microscope. *Engineering* 146, 474-475 (October, 1938).
 A new electron microscope. *J. App. Phys.* 11, 629-630 (1940).
 Polystyrene aids electron microscope. *Mod. Plastics* 21, 124-125, 160 (December, 1943).
 Technique for the electron microscopic examination of encapsulated bacteria. *Paper Tr. J.* 113, 33 (October, 1941).
 Inside rubber. *Tech. Rev.* 46, 479-481 (1944).

PATENTS

Ardenne, M. von. Magnetic objective for electron microscopes. U.S. 2,243,403. May 27, 1941.
 Ardenne, M. von. Electron scanning microscope. U.S. 2,253,542. Aug. 26, 1941.
 Ardenne, M. von. Ultraviolet or electronic microscope. U.S. 2,273,235. Feb. 17, 1942.
 Ardenne, M. von. Arrangement for controlling the anode potential and the lens currents of magnetic electron microscopes. Ger. 718,988. March 5, 1942.
 Ardenne, M. von. Electron microscope. Ger. 732,791. Feb. 11, 1943.
 Ardenne, M. von. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,356,633. Aug. 22, 1944.
 Ardenne, M. von. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,361,722. Oct. 31, 1944.
 Ardenne, M. von and Borries, B. von. Electron scanning microscope. U.S. 2,241,432. May 13, 1941.
 Boersch, H. (Assigned to General Electric Company) Electron microscope. U.S. 2,238,577. April 15, 1941.
 Boersch, H. Electron tubes for electron-diffraction apparatus or electron microscopes. Ger. 725,847. Aug. 20, 1942.
 Boersch, H. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Electron microscope for taking pictures with light and dark films. Ger. 734,736. Mar. 25, 1943.
 Boersch, H., Brüche, E., and Mahl, H. (Assigned to General Electric Company) Electron microscope. U.S. 2,259,994. Oct. 21, 1941.
 Boersch, H. and Mahl, H. (Assigned to General Electric Company) Electron Microscope. U.S. 2,249,453. July 15, 1941.
 Boersch, H. and Mahl, H. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Highly magnifying electron microscope. Ger. 729,687. Nov. 26, 1942.
 Boersch, H. and Mahl, H. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Electron microscope with two or more single lenses. Ger. 735,873. April 22, 1943. Addn. to Ger. 733,345.
 Borries, B. von. Electron microscope (Vested in the Alien Property Custodian). U.S. 2,301,987. Nov. 17, 1942.

Borries, B. von and Müller, H. O. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron optical device. U.S. 2,270,112. Jan. 13, 1942.
 Borries, B. von, Müller, H. O., and Ruska, E. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,356,551. Aug. 22, 1944.
 Borries, B. von and Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Magnetic lens for electron microscopes. U.S. 2,209,669. July 30, 1940.
 Borries, B. von and Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,268,539. Dec. 30, 1941.
 Borries, B. von and Ruska, E. Electron-optical lens (Vested in the Alien Property Custodian). U.S. 2,305,761. Dec. 22, 1942.
 Borries, B. von and Ruska, H. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Method for examining objects in an air space by means of electron microscopes. U.S. 2,256,191. Sept. 16, 1941.
 Borries, B. von, Schuchmann, H., and Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,266,717. Dec. 16, 1941.
 Brüche, E. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) High-power electron microscope. Ger. 732,690. Feb. 11, 1943.
 Brüche, E. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Electron microscope. Ger. 734,090. March 11, 1943.
 Brüche, E. and Gölz, E. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Apparatus for introducing objects into an electron microscope. Ger. 716,058. Dec. 11, 1941.
 Burkhardt, E. (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Adjusting arrangement for electron microscopes. Ger. 724,183. July 9, 1942.
 Hillier, J. (Assigned to RCA Company) Electron microscope specimen chamber. U.S. 2,272,843. Feb. 10, 1942.
 Hillier, J. (Assigned to RCA Company) Projection lens for an electron microscope. U.S. 2,323,328. July 6, 1943.
 Hillier, J. (Assigned to RCA Company) Scanning microscope. U.S. 2,330,888. Oct. 5, 1943.
 Hillier, J. (Assigned to RCA Company) Electron microscope. U.S. 2,354,263. July 25, 1944.
 Knoll, M., Houtermans, F. G., and Schulze, W. Electron microscope. U.S. 2,131,536. Sept. 27, 1938.
 Krause, F. Electron microscope vacuum system. U.S. 2,266,218. Dec. 16, 1941.
 Krause, F. (Assigned to Gewerkschaft Mathias Stinnes) Device for introducing objects into an electron microscope. Ger. 719,448. March 12, 1942. Addn. to Ger. 708,778.
 Krause, F. Preparation of specimens for examination with an electron microscope. Ger. 720,213. April 2, 1942.
 Krause, F. Method for preparing specimens of cellular material for studying in an electron microscope. Ger. 720,949. April 23, 1942. Addn. to Ger. 720,213.

- Krause, K. Device for introducing objects into an electron microscope. Ger. 708,778. June 19, 1941.
- Krause, K. Means for mounting objects in electron microscopes. U.S. 2,264,210. Nov. 25, 1941.
- Mahl, H., Pendzich, A., and Stendel, E. (Assigned to General Electric Company) Electron microscope. U.S. 2,260,041. Oct. 21, 1941.
- Marton, L. (Assigned to RCA Company) Electron microscope. U.S. 2,220,973. Nov. 12, 1940.
- Marton, L. (Assigned to RCA Company) Electronic and light microscope. U.S. 2,233,286. Feb. 25, 1941.
- Marton, L. (Assigned to RCA Company) Shielded electron microscope. U.S. 2,301,303. Nov. 10, 1942.
- Marton, L. (Assigned to RCA Company) Electron microscope. U.S. 2,347,328. Apr. 25, 1944.
- Ploke, M. (Assigned to Zeiss Ikon Aktiengesellschaft) Method of electron-microscopically investigating subjects. U.S. 2,219,113. Oct. 22, 1940.
- Rajchman, J. A. (Assigned to RCA Company) Method of focusing electron microscopes. U.S. 2,348,031. May 2, 1944.
- Rajchman, J. A. and Snyder, R. L. (Assigned to RCA Company) Electron optical instrument. U.S. 2,372,450. Mar. 27, 1945.
- Ramberg, E. G. (Assigned to RCA Company) Art of electron microscopy. U.S. 2,347,965. May 2, 1944.
- Ramberg, E. G. and Morton, G. A. (Assigned to RCA Company) Electron microscope. U.S. 2,271,990. Feb. 3, 1942.
- Ramo, S. (Assigned to General Electric Company) Electron microscope. U.S. 2,281,325. Apr. 28, 1942.
- Ramo, S. (Assigned to General Electric Company) Electron microscope. U.S. 2,363,359. Nov. 21, 1944.
- Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Shielded electron microscope. U.S. 2,234,281. March 11, 1941.
- Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,266,082. Dec. 16, 1941.
- Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,272,353. Feb. 10, 1942.
- Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,274,215. Feb. 24, 1942.
- Ruska, E. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,301,975. Nov. 17, 1942.
- Ruska, E. (Assigned to Siemens & Halske Aktiengesellschaft-Gesellschaft) Electron microscope. Ger. 734,791. March 25, 1943.
- Ruska, E., Hentschel, E., and Schuchmann, H. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,305,458. Dec. 15, 1942.
- Ruska, E. and Ruska, H. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,267,137. Dec. 23, 1941.
- Ruska, E. and Schuchmann, H. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,277,024. March 17, 1942.
- Ruska, E., Schuchmann, H., Ruska, H., and Müller, H. O. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,370,373. Feb. 27, 1945.
- Ruska, H. (Assigned to Siemens & Halske Aktiengesellschaft-Gesellschaft) Mount films for electron microscope. Ger. 709,733. July 17, 1941.
- Ruska, H. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Method of focusing electron microscopes. U.S. 2,256,171. Sept. 16, 1941.
- Schuchmann, H. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,215,979. Sept. 24, 1940.
- Schuchmann, H., Borries B. von, and Ruska, H. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,305,459. Dec. 15, 1942.
- Schuchmann, H. and Neubert, U. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,247,524. July 1, 1941.
- Snyder, R. L., Jr. (Assigned to RCA Company) Scanning type of electron microscope. U.S. 2,330,930. Oct. 5, 1943.
- Snyder, R. L., Jr. (Assigned to RCA Company) Scanning type of electron microscope. U.S. 2,348,030. May 2, 1944.
- Sukumlyn, T. W. Electron microscope system. U.S. 2,243,362. May 27, 1941.
- Young, C. J. (Assigned to RCA Company) Electron microscope object chamber. U.S. 2,347,348. Apr. 25, 1944.
- Young, C. J. (Assigned to RCA Company) Electron microscope viewing chamber. U.S. 2,356,963. Aug. 29, 1944.

No Authors Given

- (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscopes. Brit. 517,704. Feb. 6, 1940.
- (Assigned to Allgemeine Elektrizitäts-Gesellschaft) Electron-optic method for investigating mixtures of crystals. Ger. 713,740. Oct. 16, 1941.
- (Assigned to Verwaltungs-Gesellschaft m. b. H.) Electron ultramicroscope. Fr. 849,648. Nov. 28, 1940.

Contributed Original Research

Electronic Spectroscopy

GEORGE C. SZIKLAI AND ALFRED C. SCHROEDER

RCA Laboratories, Princeton, New Jersey

(Received April 12, 1946)

According to Einstein, the electrons leave a photoelectric surface with an initial velocity proportional to the frequency of the impinging radiation. Consequently, when a sawtooth potential is applied to a photoelectric cell, illuminated by a monochromatic light, the current flow starts at a threshold potential corresponding to the wavelength of the light, and then gradually increases more or less linearly toward a saturation value. When light of assorted wave-lengths is applied, the total current is the summation of the currents for the individual wave-lengths, each of which currents has its threshold at a different potential. By double differentiation an indication of the spectral distribution can be observed directly on an oscilloscope which is deflected synchronously with the potential applied to the photoelectric cell. The accuracy of the

indication is limited by the indefiniteness of the threshold values and non-linearity of the currents above the threshold potentials in available photoelectric cells. With S4 commercial surfaces, filters of the three primary colors can be easily differentiated, but with improved surfaces considerably better resolutions may be obtained and further improvement may be had by lowering the temperature of the photoelectric surface. The method lends itself readily to color matching, since two similar devices supplying their signal with opposite polarity will give zero output when the two colors match, while it provides a direct indication of the color and magnitude of mismatch when a signal is obtained. The output signal can be utilized for relays or other control devices in industrial processes.

INTRODUCTION

FIRST we may define spectroscopy, according to Sir Arthur Schuster,¹ as an investigation of radiations with respect to the wave-lengths of their constituent parts.

Spectroscopy, as a branch of physics, came into prominence after the spectroscope and its applications were introduced by Kirchoff and Bunsen in 1859, almost two centuries after Newton discovered that a prism breaks white light into its components. The optical spectroscope of Kirchoff and Bunsen was improved in many ways for greater convenience (such as the Brown, Amici, and Kessler direct line spectroscopes with compound prisms and by the use of gratings) and since 1919 through the addition of photo-cells for quantitative recording.² However, these instruments were still using purely optical principles, namely, refraction and diffraction, for the analysis of the radiation.

The present scheme uses electronic properties for the analysis and it borrows its principle from

the Einstein's law of photoelectric emission, and thus we call the instrument electronic.

THE PRINCIPLE OF OPERATION

The photoelectric equation of Einstein³ postulates that the initial kinetic energy of a photoelectron is a linear function of the frequency of the liberating radiation according to the relation of

$$mv^2/2 = h\nu - \phi,$$

where the left side represents the kinetic energy, h is Planck's constant, ν is the frequency of the radiation, and ϕ is the work function. The kinetic energy is also equal to the product of the electronic charge and the accelerating voltage; thus we may write

$$eV = h\nu - \phi,$$

where e is the charge and V is the accelerating potential required to impart to an electron the velocity V when the electron started in vacuum from rest. In the last equation all the quantities except V and ν are constant; thus we may see

¹ A lecture delivered at Royal Institution in 1882.

² K. S. Gibson *et al*, Nat. Sci. Pap. Bur. Stand. No. 349, 1919.

³ Ann. d. Physik 17, 132 (1905).

that the potential at which electron flow starts or the threshold potential is a direct function of the frequency of the radiation of a given surface.

According to the Fermi-Dirac distribution of the energy among the conduction electrons, as

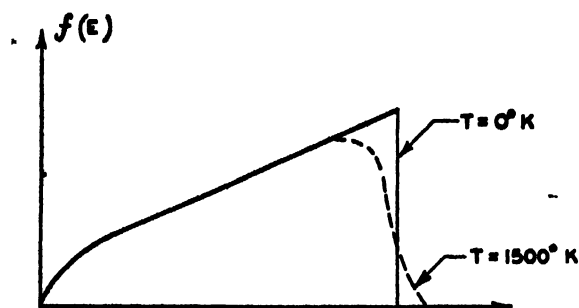


FIG. 1. The Fermi-Dirac energy distribution curve.

shown in Fig. 1, the number of electrons per unit energy is zero above a certain energy level which is lower than the work function of the metal. Figure 1 shows the distribution at 0°K and at 1500°K where the rounding off, approaching Maxwellian distribution, is caused by thermionic effect. The curve at room temperature is between the two curves shown.

Applying this information to the current in a photo-cell which is connected to a variable retarding potential, we find that if certain wavelength radiation is applied on the surface and the retarding potential is reduced to the threshold potential current flow begins. As the retarding potential is reduced further, the current increases almost linearly until saturation occurs. Figure 2 shows the effect of the radiation of two discrete bands on the current in a photo-cell as the stopping or retarding potential is reduced. The first dotted line of curve *B* shows the current produced by the higher frequency radiation, while the dotted line on the right corresponds to the lower frequency. The solid line indicates the current vs. stopping potential relation when both radiations are present. The curvature at the foot of each of the curves is caused by thermionic emission, and in some cases by non-definite work function or threshold. Such is the case with compound or non-uniform surfaces.

If curve *B* is differentiated, with respect to the stopping potential, we obtain curve *C*. A

second differentiation yields curve *A*, again reproducing the original spectra.

THE CIRCUIT

Figure 3 shows the simplest form of an apparatus performing the above described functions. A periodically variable potential (preferably sawtooth shaped) is supplied to a suitable photoelectric cell (1) by a source (2). The center potential is established by a battery or other d.c. source (3). The variation of the current changes the potential across a resistor (4) in the circuit. A differentiating circuit consisting of a capacitor (5) and a resistor (6) supplies the differentiated signal to the grid of a vacuum tube amplifier (7). This amplifier is followed by a similar differentiating circuit-feeding amplifier (8). The amplified twice-differentiated signal is then applied to the vertical deflection system of a cathode-ray oscilloscope (9) which is deflected horizontally with the same periodicity as the potential across the cell (1) is varied.

Figure 4 shows a simple circuit for differentiation with two photoelectric cells (10) and (11) connected in series and biased with the aid of slightly different d.c. sources (12) and (13) at different center potentials. The periodical wave is applied by source (2') through a center tapped

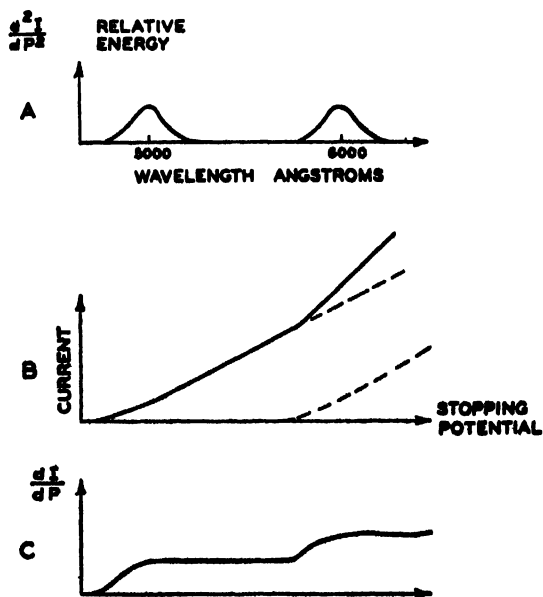


FIG. 2. Graphical representation of the photo-current differentiation.

transformer (14), and the differentiated output is obtained across resistor (4'). This output then is differentiated again by capacitor (5') and resistor (6'), amplified by amplifier (8') and applied to the oscilloscope (9') similarly to the way shown in Fig. 1.

Figure 5 shows the manner in which differentiation is taking place by the circuit shown in Fig. 4. Curve *A* shows the voltage developed by one photoelectric cell when it is exposed to a light source of red and blue content as shown by curve *D*. Curve *B* is the same as *A* with opposite polarity, except it is delayed by the above-mentioned bias difference. The resultant potential (the sum of the two potentials) that appears across resistor (4') is shown as curve *C*, which is substantially the differential of curve *A*. If signal *C* is differentiated the resultant potential will be similar to the curve *D*.

A special circuit is shown in Fig. 6, which is similar to Fig. 4, except that one of the cells observes the light to be tested, while the other

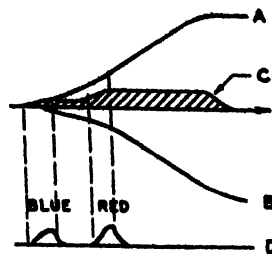


FIG. 5. Differentiation by push-pull action.

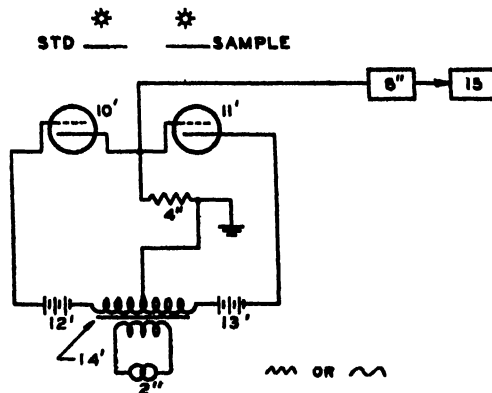


FIG. 6. Color matching diagram

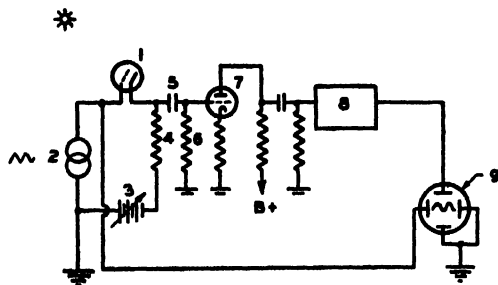


FIG. 3. Simple circuit with double R-C differentiation

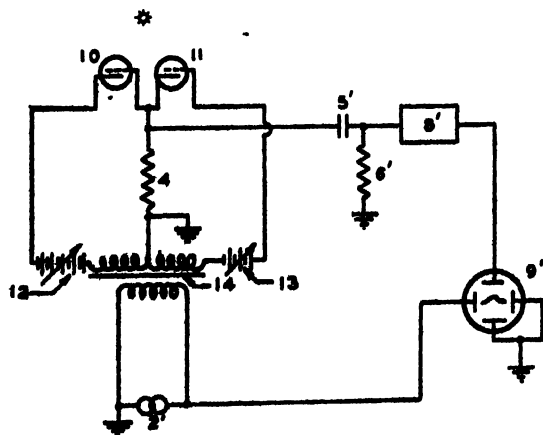


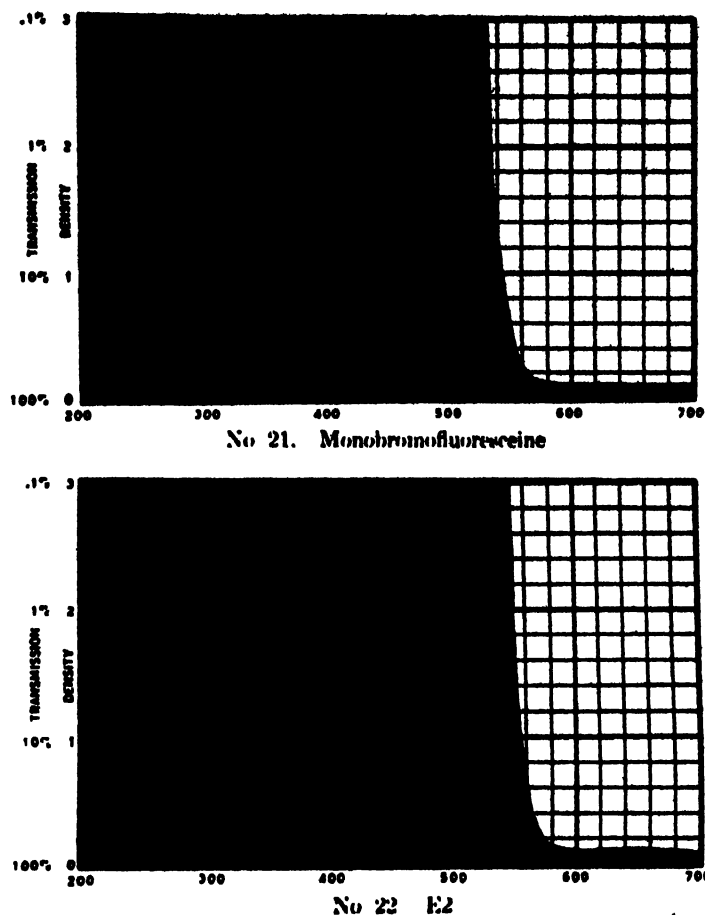
FIG. 4. Circuit with differentiation in push-pull cells.



FIG. 7. The experimental unit.

provides a signal from a standard to be matched. The two bias potentials (12') and (13') are equal, and thus if the two light sources (or reflectors) have the same spectral distribution, the currents through cells (10') and (11') are the same, and the output signal across resistor (4') is zero. The amplifier (8'') provides an actuating signal to relay 15 only if the two colors are not matched. This circuit finds a particular application in

FIG. 10. Published density curves of Wratten No. 21 and 22 filters.



Another step to obtain a sharp threshold is the use of pure metal surfaces eliminating the non-uniformity of the work function over the surface.

Frohlich⁵ has obtained extremely sharp thresholds from thin alkali metal films presumably due to eliminating the effects of absorption of light and of electrons within the film.

⁵ Ann. d. Physik 7, 103 (1930).

While the present results are adequate for indications in industrial applications, further refinements may provide an instrument with greater precision.

The simplicity of the instrument and its convenient use due to its instantaneous application ought to be an incentive for further work in making the required refinements.

Calculated Frequency Spectrum of the Shot Noise from a Photo-Multiplier Tube*

R. D. SARD

Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received April 4, 1946)

The mean square noise current per unit frequency range as a function of frequency, or "power spectrum," of the shot noise produced by a secondary-emission multiplier tube is calculated. A general expression for it, (3), is set up and then applied to the RCA 931 family of photo-multiplier tubes. These tubes are of interest as relatively intense sources of random noise of constant intensity up to high frequencies. When an electron enters the collecting space, it initiates a pulse of current in the anode circuit. In the 931 operated with 100 volts per stage, the pulse has a total duration of about 6×10^{-10} sec and a half-width of about 1.5×10^{-10} sec. The electrons descended from the same photon do not enter the collecting space simultaneously but are spread out over a time of the order of 6×10^{-10} sec. Because of the non-zero duration of the pulses and the spread in their time of occurrence, the

noise spectrum is essentially flat only up to several hundred Mc, falling rapidly at higher frequencies. The paucity of data and the approximations made in the calculation make our curve of noise output *versus* frequency significant only as regards order of magnitude and general trend. It does seem safe to predict that the noise intensity ought to be constant from zero up to about 100 Mc, begin to fall off appreciably between 100 Mc and 1000 Mc, and become very weak at higher frequencies. Partial confirmation is provided by unpublished measurements of Jastram at the Radio Research Laboratory, Harvard University, which show that at 30 Mc the noise output is still about the same as in the range 0 to 5 Mc. There is also some experimental indication that the output is about the same at 100 Mc and that it is much smaller at higher frequencies

1. INTRODUCTION AND SUMMARY

THE subject of this paper is the frequency spectrum of the noise current in a vacuum secondary-emission multiplier tube. Under normal operating conditions, the fluctuations in collector potential associated with the noise fluctuations are small compared with the applied constant accelerating potential differences, and it is therefore legitimate to consider the noise current as independent of the external load. After the mean square noise current per unit frequency interval has been calculated on this assumption, the mean square noise voltage can be obtained from it by multiplying with the square of the absolute value of the load impedance and integrating over all frequencies.

We derive first a very general expression (Eq. (3)) for the mean square noise current per unit frequency interval for any type of multiplier tube, and then evaluate it for the RCA 931 family of photo-multiplier tubes. Approximations are made in the evaluation that are of a crudity commensurate with the meagerness of the available data on the properties of the secondary-emitting surfaces and the electron trajectories in

these tubes, and with the variations in behavior from tube to tube. For this reason, the numerical results are significant for frequencies above several hundred megacycles only as regards order of magnitude and general trend. The calculation may, nevertheless, be of interest to physicists and radio engineers, not only because of the physical ideas involved, but also because of the possibility of using a tube like the 931 as a laboratory noise source for the study of receiver sensitivity. Because of its high gain, as explained below, this tube gives a large mean square noise current for a given average value of the anode current. And the noise is essentially white, i.e. its frequency spectrum has a constant intensity, from zero to at least one hundred megacycles.

Shot noise in a tube of any kind is caused by the fact that charge is brought to the collector in discrete lumps whose times of arrival are not correlated with each other. In a saturated diode, the individual electrons moving from cathode to anode constitute the lumps. Each such transit electron moves independently of the others, producing its independent transient current in the plate circuit. The piling-up of these transients produces the plate current. Because the transients occur at random, the plate current undergoes random fluctuations about its average value. In a secondary-emission multiplier tube,

* This paper is based on work done for the Office of Scientific Research and Development under Contract No. OSM-471 with the President and Fellows of Harvard

the phenomenon of secondary electron emission is used to magnify the primary current. By the use of a number of secondary-emitting stages in cascade, one primary electron is made to produce a large number of electrons in the last stage of multiplication. Thus, in the 931, which has nine stages of multiplication, one photon incident on the photo-sensitive cathode liberates a photo-electron which then produces hundreds of thousands of electrons in the ninth generation. These electrons are collected by the anode at almost the same instant. The independent lumps in which charge reaches the anode are, therefore, hundreds of thousands of times larger than in the case of a saturated diode. The effect is equivalent to an increase of the electronic charge by this factor in the formula for the shot noise of a saturated diode.

As an electron approaches the anode it induces a pulse of current in the external circuit to the anode. We shall refer to a current pulse due to an individual electron as an *elementary pulse*. The sum of the elementary pulses due to all the electrons descended from the same primary will be called a *burst*. As the primary photoelectrons occur at random, the bursts that they produce also occur at random. The bursts are the *random elements*, or lumps, which pile up randomly to form the anode current.

It is seen that in general the theory of shot noise falls naturally into two parts: a physical part concerned with the amplitude and shape in time of the random elements (in our problem, the bursts) and a statistical part concerned with the various mean values resulting from the random superposition of these elements.

We have to calculate the noise *power spectrum*, the mean square noise current per unit frequency interval as a function of frequency. The statistical part of the problem is disposed of by the well-known theorem¹ that the average power spectrum of the sum of a number of random elements is the sum of the power spectra of the individual elements (by "average" is meant the average over an ensemble of independent intervals, in each of which a Fourier analysis is made; the result is approached asymptotically by the spectrum of a single interval as its duration

increases). We need only to calculate, then, the power spectra of the individual bursts. While considerable variations in the amplitudes of the bursts may be expected (and are, indeed, taken into account in our calculation), their shapes as functions of time should be nearly the same. We therefore calculate the power spectra of the bursts on the assumption that the bursts all have one typical shape.

The shape of a burst is determined by the shapes of the constituent elementary pulses and by their distribution in time. If the spread in time of occurrence of the elementary pulses is small compared to their duration, the elementary pulses simply pile up right on top of one another, and the burst shape is determined by the elementary pulse shapes. If the spread in time of occurrence is large compared to the duration of the elementary pulses (but still small enough to give considerable overlapping of the elementary pulses), the shapes of the elementary pulses are washed out, and the burst has the shape of the distribution function of time of occurrence of the elementary pulses. If the durations of the elementary pulses and the spread in their times of occurrence are of the same order of magnitude, both must be taken into account. It turns out that in the 931 a typical elementary pulse has a total duration of about 6×10^{-10} sec. and a duration between half-amplitude points of about 1.5×10^{-10} sec. The distribution curve for time of occurrence of the elementary pulses has a width between half-amplitude points of about 6×10^{-10} sec. Hence, both the duration of the elementary pulses and their spread in time of occurrence must be taken into account.

The non-zero duration of the elementary pulses is caused by the finite velocities with which the electrons approach the anode. As an electron approaches the anode, it induces on it an increasing amount of positive charge, which is supplied by a current in the external circuit. This current is equal to the rate of change of the charge induced on the anode, being a pulse of duration equal to the time interval during which an appreciable fraction of the electron's lines of force end on the anode. It is our elementary pulse. To calculate its shape, we need to determine: (1), the position and velocity of the electron as a function of the time; and (2), the

¹ Rayleigh, Phil. Mag. 28, 460 (1889).

amount of charge induced on the anode by the electron as a function of its position. The first problem is the dynamical one of the motion of an electron in a given field of force; the second is an electrostatic one which can be solved in certain cases by the method of images.³

The spread in time of occurrence can be owing to: (1) fluctuating time lags in secondary emission, (2) spread in initial velocities of the secondary electrons. The spread in emission velocities has two results. The secondaries liberated by a particular impinging electron have different times of flight to the next electrode. Secondly, they reach it at different points, so producing a spatial spreading of the cascade. This spatial spreading entails variations in times of flight for different branches of the cascade. Reasons are given in Section III for believing that in the 931 the main cause of spread in time of occurrence of the elementary pulses is the spread in emission velocities as it affects the time of flight from a particular point of one electrode to the next electrode. The other effects just mentioned should involve smaller time spreads, and should, therefore, only affect the noise spectrum at higher frequencies.

The time elapsed between the arrival of the primary electron at the first multiplication electrode and the beginning of an elementary pulse is the sum of eight independently fluctuating inter-stage transit times. The most probable total elapsed time is therefore the sum of the most probable inter-stage transit times; and the mean square deviation of the total elapsed time is the sum of the mean square deviations of the inter-stage transit times. As eight is a large number from the point of view of the central limits theorem of statistics, the distribution curve of time of occurrence of the elementary pulses will be Gaussian.

³ This paragraph appears to contradict the atomicity of electricity, as it implies that the induced charge varies continuously from zero to one electron during the flight of the transit electron. Professor Van Vleck has pointed out to the author that the paradox disappears when one considers the atomic nature of the external circuit. When we speak of the induced charge on a body, we are really referring to its electric polarization; the induced charge is the result of the displacement of the metallic conduction electrons. This polarization is a continuous variable. The quantum-mechanical justification for the use of electrostatic image theory when the inducing charge is a single electron has been given by J. Bardeen (Phys. Rev. 58, 727 (1940)).

In our calculation of the burst spectrum, we assume that the elementary pulses constituting a burst all have essentially the same shape. We then use a well-known theorem in Fourier analysis⁴ which, applied here, states that the spectrum of the burst is the product of the spectrum of the elementary pulse and the spectrum of the distribution curve of time of occurrence of the elementary pulses. This saves us the labor of computing the burst shape itself. We need only calculate the mean elementary pulse and the distribution curve of time of occurrence, find the Fourier transform of each, and then take their products.

The weakest point in our calculation for the 931 would seem to lie in the calculation of the spread in inter-stage transit time. No data were available on the secondary emission velocities from the surfaces actually used in these tubes, and very little was known about the electron trajectories. In order to get some idea of the orders of magnitude, we have used Malter's published data⁴ on normal emission velocities from a silver-magnesium surface similar to that used in the 931's, applying them to the case of plane parallel electrodes. The result of the use of this crude approximation and of other less dubious ones, and of the physical variations from tube to tube, should make our results significant only as to order of magnitude and general trend. The initial falling off in noise output with rising frequency should, however, be given fairly accurately; and it does seem safe to predict that the noise spectrum should be essentially flat up to about 100 Mc and start to fall off appreciably in the range of 100 to 1000 Mc. Partial confirmation is provided by unpublished measurements of Jastram at the Radio Research Laboratory, Harvard University, which show that at 30 Mc the noise output is still about the same as in the range 0 to 5 Mc. There is also some experimental indication that the output is about the same at 100 Mc and that it is much smaller at higher frequencies.

⁴ E.g., G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Applications* (Bell Telephone System Publications, New York, 1931), Monograph B-584, formula 202.

⁴ L. Malter, Proc. I.R.E. 29, 587 (1941).

II. GENERAL THEORY OF THE SHOT NOISE IN A MULTIPLIER TUBE

Let the elementary pulse due to the i th electron have a shape represented by the function $\Delta_i(t-t_i)$, where t_i is the time at which the pulse ends (electron reaches anode). If we normalize Δ_i to unit area, the elementary pulse is $e\Delta_i(t-t_i)$ where e is the electronic charge. The total anode current is

$$i(t) = \sum_i e\Delta_i(t-t_i). \quad (1)$$

This is the current to the anode through the external circuit, which includes internal capacitances.

We can group together the elementary pulses belonging to the same burst (i.e., descended from the same primary), and write

$$i(t) = \sum_k eM_k B_k(t). \quad (2)$$

The index k refers to the burst. M_k is the number of electrons in the burst; $B_k(t)$ is the shape of the k th burst, which we have normalized to unit area.

Expanding the current in a Fourier series of fundamental period T (long enough so that the number of harmonics in a small frequency range $\Delta\nu$ is proportional to $\Delta\nu$), we obtain

$$\langle P(\nu) \rangle_n = 2e \langle i(t) \rangle_n \frac{\langle M^2 \rangle_n}{\langle M \rangle_n} \cdot \langle |TB(\nu)|^2 \rangle_n, \quad (3)$$

where $B(\nu)$ is the Fourier transform of the burst shape $B(t)$:

$$B(\nu) = \frac{1}{T} \int_0^T B(t) \exp(-i2\pi\nu t) dt, \quad (3a)$$

and $P(\nu)$ is the mean square noise current per unit frequency range. The derivation involves a verification of Rayleigh's result.¹ No physical assumption about the bursts is involved other than that they occur at random; they need not be similar in shape. Formulas (3) and (3a) are the basis for the calculations of this paper.

It is interesting to note from (3) that fluctuations in the multiplication factor increase the noise power² because of the factor $(\langle M^2 \rangle_n / \langle M \rangle_n)$.

¹ Ziegler (Physica 3, 307 (1936)) has pointed out how measurement of the mean square noise current from a one-stage multiplier gives the mean square number of

The effective size of the lumps of charge is not $e\langle M \rangle_n$ but rather $(e\langle M^2 \rangle_n / \langle M \rangle_n)$.

For frequencies so low that the period is large compared to the burst duration, $B_k(t)$ can be replaced by a δ function, and

$$B_k(\nu) = \frac{1}{T} \exp(-i2\pi\nu t_k); \quad (4)$$

where t_k is the time of occurrence of the burst. (3) becomes

$$\langle P(\nu) \rangle_n = 2e \langle i(t) \rangle_n \frac{\langle M^2 \rangle_n}{\langle M \rangle_n}. \quad (4a)$$

This special case of (3) has been derived by Shockley and Pierce.³ For higher frequencies, the shape of the burst becomes important in determining $B_k(\nu)$. Finally, when the frequency becomes so high that a large number of oscillations take place in a time in which $B_k(t)$ changes but little, $B_k(\nu)$ becomes small, and approaches zero. In general, then, the noise spectrum is flat at zero frequency and approaches zero at high frequency (inversely as the square of the frequency), with an intermediate region in which its behavior is determined by the burst shape.

In practice the fluctuations in the potential of the anode relative to the preceding stage are small compared to the applied d.c. potential differences. This justifies our calculation of a given noise current independent of the load. The measured noise voltage will depend on the impedance of the load, which, as already remarked, includes the coupling of the anode to the other electrodes and leads. Calling the total impedance $Z(\nu)$, we have for the mean square voltage across the load

$$\langle v^2 \rangle_n = 2T \int_{-\infty}^{\infty} |Z(\nu)|^2 \cdot \langle |I(\nu)|^2 \rangle_n \cdot d\nu, \quad (5)$$

where $I(\nu)$ is the transform of $i(t)$. This gives for

secondary electrons liberated by one primary. It may be worth remarking that measurement of the higher moments of the output noise would give the higher moments of the distribution function for the number of secondaries per primary. In this way, the probabilities of emission of 0, 1, 2, 3, ... secondaries per primary could be determined.
³ W. S. Shockley and J. R. Pierce, Proc. I.R.E. 26, 321 (1938).

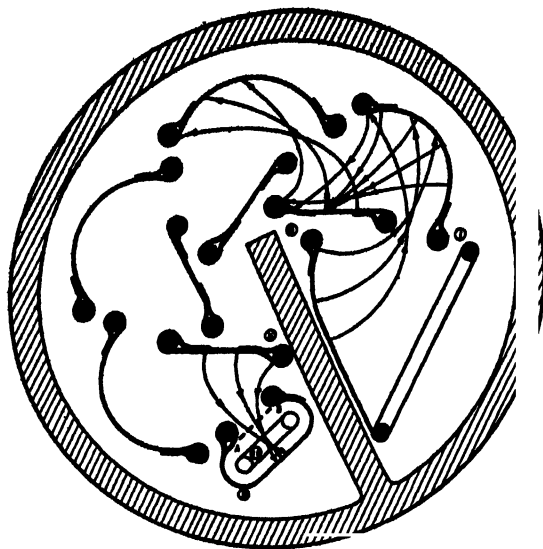


FIG. 1. Internal structure of the RCA 931 photo-multiplier tube. The photoelectrons from (0) are drawn over to (1), the first dynode. The secondary electrons leaving (1) are drawn to the next dynode, and so on through nine stages of multiplication, the last occurring at dynode (9). (10) is a double grid serving as the collector. It is almost entirely enclosed by (9). The dashed line AB represents the estimated position of an equipotential joining on to (9). It is assumed in our calculation that only when an electron is in the space enclosed by (9) and AB does an appreciable fraction of its lines of force end on (10).

the mean square fluctuation voltage

$$\langle v^2 \rangle_n - \langle v \rangle_n^2 = 2e \langle i(t) \rangle_n \frac{\langle M^2 \rangle_n}{\langle M \rangle_n} \times \int_{-\infty}^{\infty} \langle |TB(\nu)|^2 \rangle_n \cdot |Z(\nu)|^2 \cdot d\nu. \quad (5a)$$

When the impedance differs from zero only in the low frequency region where (4a) applies,

$$\langle v^2 \rangle_n - \langle v \rangle_n^2 = 2e \langle i(t) \rangle_n \frac{\langle M^2 \rangle_n}{\langle M \rangle_n} \int_{-\infty}^{\infty} |Z(\nu)|^2 \cdot d\nu. \quad (5b)$$

Setting $M = 1$ gives the well-known expressions for the saturated diode.

Our problem is seen to be that of the calculation of $B_k(\nu)$. Now, in the k th burst, the time of arrival, t_i , of each elementary pulse is distributed about a mean value $\langle t_i \rangle_k$ according to some distribution function for the spread in time of arrival, $f_k(t_i)$, which we normalize to unit area. Since M is a very large number (over the range

of 931's and usual operating conditions it lies between 3×10^4 and 3×10^6) we may write

$$B_k(t) = \int_{t_i = -\infty}^t \Delta_k(t - t_i) \cdot f_k(t_i) dt_i, \quad (6)$$

on condition that it is permissible to replace the Δ_i 's by Δ_k , a mean shape for the elementary pulses making up the k th burst. If use of this mean shape is legitimate, we may also regard t_i in (6) as the time at which the elementary pulse starts, and $f_k(t_i)$ as the distribution function for time of arrival at the starting point.

The applicability of the assumption in the case of the 931 is discussed in Section III.

Since $B_k(t)$ has the form (6), we can apply a well-known theorem in Fourier analysis⁸ to obtain: *The Fourier transform $B_k(\nu)$ of the burst shape is the product of the Fourier transform of the elementary pulse shape and the Fourier transform of the distribution curve for time occurrence of the elementary pulses.* Thus we can consider separately the effects of finite velocity of the electrons and spread in their time of arrival.

III. APPLICATION TO THE RCA 931

The RCA 931⁷ is an electrostatic multiplier with nine stages of multiplication. Figure 1 shows the structure of the tube, and some of the electron trajectories.⁸ Electrode (0) is the photo-sensitive cathode. The photoelectrons are drawn to electrode (1), the first multiplying electrode or "dynode," where the process of multiplication starts. It continues through nine stages. The arrangement of the ninth dynode and the anode (electrode (10)) should be noted. The anode is a double grid almost entirely enclosed by (9). The electrons from (8) follow the trajectories shown, passing through the imaginary surface AB (an equipotential joining on to (9)), then through the Faraday cage bounded by the two

⁷ RCA has replaced the 931 by the 931-A, which is the same tube made to closer tolerances, and has also put on the market other versions of the tube with the same geometry but different sensitivity and power rating. The 1P21 also uses the "S4" surface, but has higher gain and lower power rating. The 1P22 uses the "S8" surface, the 1P28 the "S5." See *RCA Technical Data Sheets* for details of these tubes. The calculations made in this report should apply within the indicated limits to the whole family.

⁸ I am indebted to Dr. J. Rajchman of the RCA Laboratories for permission to use this figure.

sides of (10), and thence to (9). This produces a bi-directional current pulse in the anode circuit, positive while the electrons approach (10), zero while they are inside (10), and negative while they are moving away. As the area under it totals zero, it contributes only to the high frequency noise. At (9) the electrons multiply, and the secondaries are drawn back to (10), giving a positive pulse on the heels of the weaker two-sided one. The pulse shape, calculated as described below, is shown in Fig. 2; in this figure m is the multiplication at (9).

Because of the electrode structure, the model used in the derivation of II—where the elementary pulse Δ does not start till all the multiplication has taken place—does not apply exactly. But a slight broadening of the assumption made in Eq. (6) covers the situation. We can regard $f_k(t_s)$ as the distribution function for time of arrival at AB , t_s being that time and $\Delta_k(t-t_s)$ as the average elementary pulse $ABCDEFGHIJ$ of Fig. 2, with m set equal to the average multiplication at (9). It is almost certain that the neglected variations in Δ —due now to

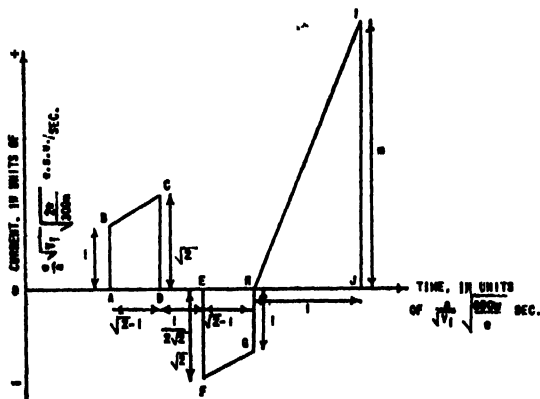


FIG. 2. The elementary current pulse in the external circuit due to the entrance of a single electron into the collecting space. With the inter-stage voltage $V_1=100$ and the distance a (see Fig. 3) 0.085 cm, the unit of time is 2.9×10^{-10} sec. and the unit of current is 1.1×10^{-9} amp. At A, when the electron crosses the imaginary surface AB of Fig. 1, an appreciable fraction of the electron's lines of force begins to terminate on the anode (10). As the electron is accelerated toward the anode, the induced current increases (BC). It falls to zero (DE) when the electron gets into the Faraday cage formed by the two sides of (10). On emerging from the other side of (10), the electron induces a current in the opposite direction, which decreases as the electron is slowed down (FG). On reaching (9) the electron is brought to rest (GH) and m secondaries are produced (with $V_1=100$, m is around 4). These secondaries are accelerated toward (10), giving the forward pulse HJJ , which is terminated when they reach (10) and are collected.

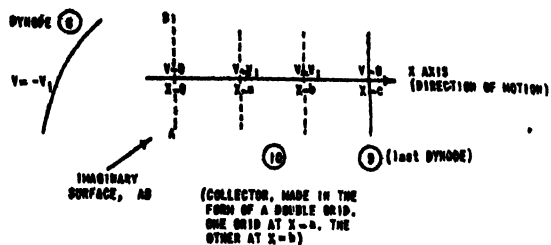


FIG. 3. Equipotentials in the collecting space. The electrons move along the X axis. After crossing AB ($x=0$) they are accelerated toward the anode ($x=a$), coast through it to its other side ($x=b$), and are decelerated as they approach the last dynode (9) at $x=c$. They are stopped there, and the secondary electrons produced are drawn (in the direction of decreasing x) to the anode, where they are collected.

fluctuations in m as well as to time and space spreads between AB and (10)—only begin to affect the spectrum at frequencies so high that the spread in time of arrival at AB and the duration of the average elementary pulse have already reduced the noise output appreciably. The inaccuracy introduced into our calculated spectrum is probably no greater than that caused by variations in performance from one 931 to another, the result of both being to make our results above several hundred megacycles significant only as regards order of magnitude and general trend.

Let us now calculate the average elementary pulse of Fig. 2. It would seem to be a fairly good approximation to assume that only when the electron crosses AB (Fig. 1) does it begin to induce an appreciable amount of charge on the anode. We need, then, to calculate the motion of the electron from this time till the instant it reaches (9), and the motion of its secondaries from the instant they leave (9) till they arrive at (10). It is seen from Fig. 1 that the idealization of plane parallel equipotential surfaces with motion perpendicular to them corresponds closely to the actual state of affairs. It is also a good approximation to take $a=b-a=c-b$ (see Fig. 3). In Fig. 3 the potential is measured from that of (9) taken as a reference level. Under usual conditions, space charge is negligible,⁷ and the dynamical problem is the simple one of the motion of a single electron in a uniform field.

The initial velocities of the secondary electrons are of the order of 5 volts.⁴ As this is only a small fraction of the velocity acquired in falling to the

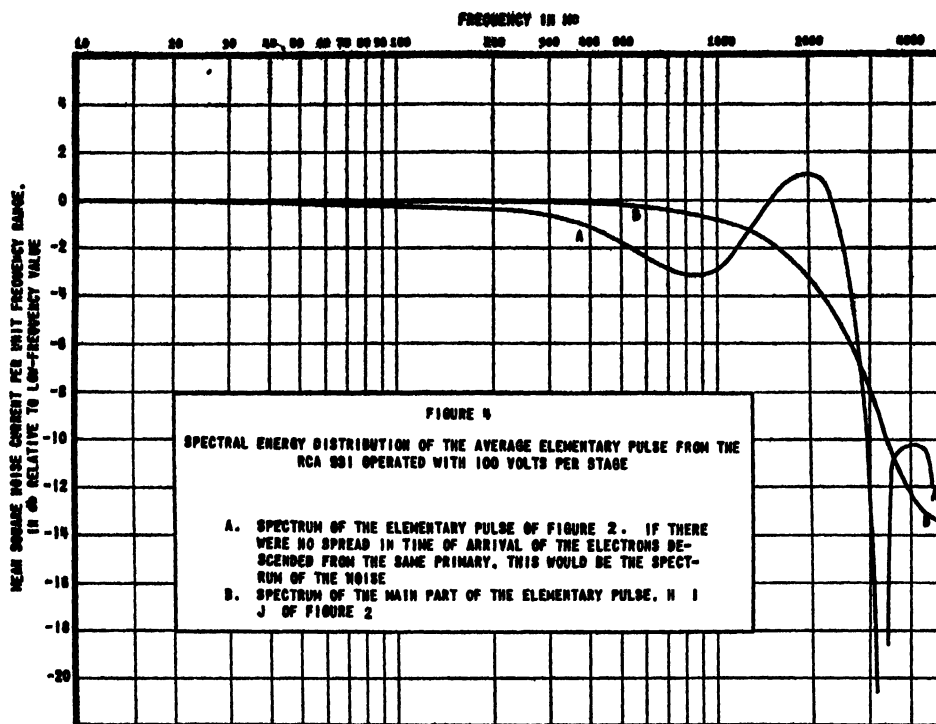


FIG. 4. Power spectrum of the average elementary pulse in the RCA 931 operated with 100 volts per stage. Curve *A* is the power spectrum of the elementary pulse shown in Fig. 2. If there were no spread in the time of arrival of the electrons descended from the same photon, this would be the spectrum of the noise. Curve *B* is drawn in for comparison. It is the spectrum of the main part of the pulse, *H I J* of Fig. 2, considered by itself. This would be the spectrum of the noise from a saturated plane parallel diode with 100 volts potential difference and electrode spacing of 0.085 cm (assuming zero emission velocity).

next electrode, we may neglect the initial velocities of the secondaries from (8) and (9) in the calculation of the average elementary pulse.

Straightforward solution of the dynamical problem gives the position and the velocity as a function of the time. At *AB* (Fig. 3) the electron has the velocity corresponding to V_1 (5.9×10^8 cm/sec. for $V_1 = 100$ volts). In going to $x = a$, its velocity increases linearly with the time, reaching a value $\sqrt{2}$ times that at *AB*. It then coasts at the same speed from $x = a$ to $x = b$. Between $x = b$ and $x = c$, the velocity decreases linearly to the value corresponding to V_1 . The secondaries made at $x = c$ are then accelerated back to $x = b$, acquiring the velocity corresponding to V_1 .

With $V_1 = 100$ volts and $a = 0.085$ cm, the total time elapsed is 6.3×10^{-10} second.

As to the electrostatic problem: the distribution of the lines of force of a point charge between two infinite parallel conducting planes has been

worked out by Maxwell in his *Treatise*.⁹ With the planes a distance d apart, the induced charge on one plane is ey/d , where e is the value of the point charge and y is its distance from the other plane. It follows that the induced current is given by e/d times the velocity.

At $x = a$, $x = b$, and $x = c$ there are conducting planes. The plane *AB* ($x = 0$) is, however, merely a mathematical surface that is no longer equipotential when an electron is in its vicinity. Our assumption that no lines of force from an electron reach 10 until it has passed *AB* is, however, equivalent to assuming that *AB* is a conducting surface, and we shall therefore make no further error by considering it as such. Applying Maxwell's result to our solution of the dynamical problem, we obtain for the current to (10) the

⁹ J. C. Maxwell, *Treatise on Electricity and Magnetism* (Oxford University Press, England, 1892), third edition, Vol. 1, p. 266.

result shown in Fig. 2. This is the average elementary pulse. With $V_1=100$ volts and $a=0.085$ cm, the unit of time is 2.9×10^{-10} sec.; the unit of current is 1.1×10^{-9} amp. m is the multiplication at 9. With 100 volts per stage, its average value is around 4; we use $m=4$ in our calculation. Note that while the pulse lasts 6.3×10^{-10} sec., the main part of it is only 2.9×10^{-10} sec. long and has a width between half-amplitude points of only half this.

Curve A of Fig. 4 is the Fourier analysis of this pulse. The ordinate is the mean square noise current per unit frequency range, relative to that at low frequency, in decibels (db). The peculiar shape of the elementary pulse accounts for the rise around 1900 Mc, where the positive and negative portions of the pulse are in resonance. The main part of the pulse, HIJ of Fig. 2, has, by itself, the spectrum shown in curve B.¹⁰

At 254 Mc ($2\pi\nu\tau=1$, where $\tau=6.3 \times 10^{-10}$ sec.), the spectrum is down less than 1 db. Though it falls to -3.2 db at 900 Mc, it does not reach -6 db till around 2800 Mc.

If the accelerating voltages between (8) and (9) and (9) and (10) were increased by the same factor, the frequency scale of Fig. 5 would be shrunk by the square root of this factor (i.e., the spectrum would be flat to higher frequencies). Note that only the last two inter-stage voltages are involved; the preceding ones have no effect on the spectrum of the elementary pulse.

We now have to consider the other major cause of frequency dependence, the spread in the time of arrival of the electrons at AB (Fig. 1). This may be because of: (1), fluctuating time delays in secondary emission and (2), differences in the initial velocities of the secondary electrons. As a consequence of (2), the secondaries liberated by a particular electron will take different times to reach the next dynode. They will also arrive there at different points; this will lead to a spatial spreading of the cascade, giving different times of flight for the different branches. Little is known about (1), other than that spread due to it in a five-stage multiplier of design similar to that of the 931 is negligible up to 500 Mc.⁴

¹⁰ This curve agrees essentially with those calculated for a saturated diode with zero emission velocity by Ballantine, J. Frank. Inst. 206, 159 (1928) and Spenke, Wiss. Veröff. Siemens [III], 16, 127 (1937).

We shall assume that it only enters at frequencies so high that we can neglect it in our calculation. The spread in emission velocities undoubtedly has a considerable effect. As just remarked, it makes itself felt directly in the time of flight from one point of a dynode to the next dynode and indirectly through the geometrical widening of the cascade. The latter effect is important when one considers the amplification of a high frequency signal,⁴ but it should have less bearing on the noise problem. In effect, we are only concerned with the time spread *within* an individual burst, originating with one electron leaving some point of the photo-cathode; its spatial spreading is due to the variations in velocity of the secondaries made in the subsequent stages. As the initial kinetic energies are small relative to the accelerating potential differences, the spread should be small in the first few stages; it should be kept within bounds in the later stages by the concentration of the beam from stage to stage.⁴ The only quantitative information that we have on this point has been supplied by Glover, Engstrom, and Pietenpol of the RCA Laboratories, who state that the spread parallel to the axis of the tube is of the order 0.5 cm in the last stage. Dispersion in this direction should have little effect on the time of arrival at the surface AB; and spread in the plane perpendicular to

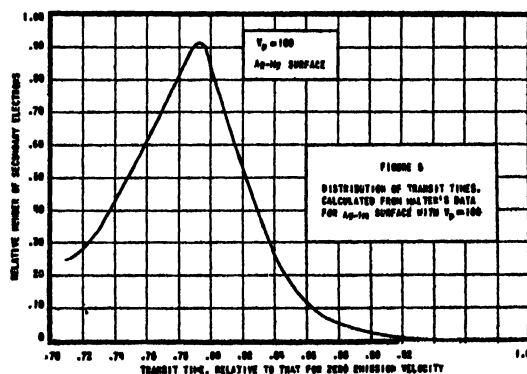


FIG. 5. Distribution of inter-stage transit time. The curve shows the relative number of secondary electrons per unit transit time as a function of the transit time. The transit time is measured in terms of that with zero emission velocity, and is therefore less than or equal to 1. The curve has been computed from Malter's measured distribution of normal emission velocities for the secondaries liberated from an Ag-Mg surface bombarded by 100-volt electrons. It has no quantitative significance for the 931, but should serve in our order of magnitude calculation.

the axis should be much less, because of focusing in this plane.

Calculation of the direct effect of spread in emission velocity on inter-stage transit time requires knowledge of the distribution in speed and direction of the secondaries, as well as of the electrostatic field between stages. No data on the velocities from the "S4," "S5," or "S8" surfaces is known to the author; also, it is clear from Fig. 1 that the electron optics of the tube is rather complicated. For the case of plane parallel electrodes, the transit time is determined simply by the normal component of the emission velocity, and Malter has published⁴ curves for the distribution of this component, for a silver-magnesium surface bombarded by 100-volt and 200-volt primaries. Figure 5 has been calculated from the curve for 100-volt primaries in his Fig. 3; it shows the number of secondaries per interval of transit time as a function of the transit time, which is measured relative to the maximum possible one (zero initial velocity). What with the directionality of secondary emission (cosine law¹¹), the focusing in the tube, and the similarity between the Ag-Mg surface and those used in the various forms of the 931, the curve of Fig. 5, should serve in a calculation of the order of magnitude of the expected effect.

The most probable transit time is seen to be 79 percent of the maximum, given, in seconds, by

$$3.37 \times 10^{-8} (d/V)^{1/2},$$

where d is the distance between the electrodes in cm and V the potential difference in volts. The half-width of the distribution curve is 10.5 percent of the most probable value; this half-width should not be too different from twice the root mean square deviation.

The over-all transit time to AB involves the sum of eight independently fluctuating inter-stage transit times (from (1) to AB). The spread in the time of arrival at AB can therefore be calculated from the statistical theorem which states that the mean square deviation of the sum of a number of independently fluctuating variables is the sum of the mean square deviations of the independent variables. Calculating the

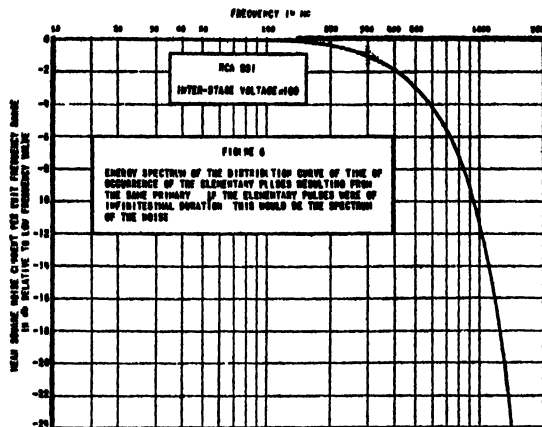


FIG. 6. "Power" spectrum of the distribution curve of time of occurrence of the elementary pulses resulting from the same photon. This is the Fourier transform of the Gaussian distribution curve for the time of occurrence of the elementary pulses constituting a single burst. If the elementary pulses were of infinitesimal duration, the curve would be the power spectrum of the noise. The curve is drawn for an inter-stage voltage of 100.

stage-to-stage spread as indicated in the preceding paragraph, and applying the theorem, we obtain 6×10^{-10} sec. for the spread in time of arrival at AB of the electrons descended from a common primary.

We note that the spread in transit time from (9) to (10) is only 2.5×10^{-11} second, which is small compared to the duration of the elementary pulse and the spread in time of arrival at AB .

As the spread in time of arrival at AB is the resultant of a number of independent fluctuations (eight) that is large from the point of view of the central limits theorem, its distribution curve is normal:

$$\exp \left(\frac{-(t-t_m)^2}{2(.425\Delta t)^2} \right),$$

where Δt is the half-width and t_m the most probable time of arrival. Its Fourier amplitude is, therefore, proportional to

$$\exp \left(\frac{-(.425)^2}{2} (2\pi\nu\Delta t)^2 \right),$$

which is plotted in Fig. 6 with $\Delta t = 6 \times 10^{-10}$ sec. At 265 Mc ($2\pi\nu\Delta t = 1$) it is down less than 1 db, but it falls very rapidly, in decibels as the square of the frequency.

¹¹ M. Lortie, *Le phénomène physique d'émission secondaire* (CSA-SFR Conférences de Documentation, Paris, May, 1942).

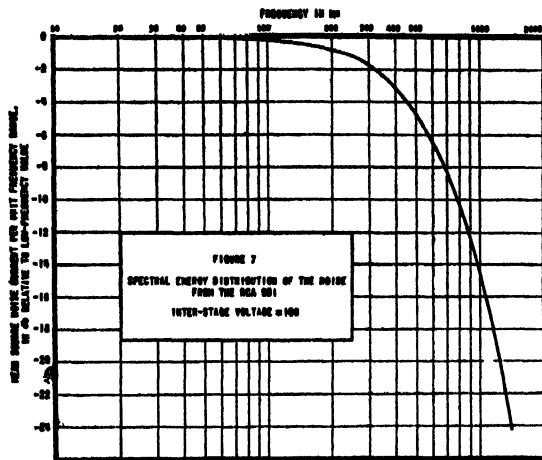


FIG. 7. Power spectrum of the noise from the RCA 931 operated with 100 volts per stage. This curve shows the mean square noise current per unit frequency range *versus* frequency. It is obtained by combining curve *A* of Fig. 4 with the curve of Fig. 6. The rapid fall in the spectrum of the distribution curve of time of arrival of the individual electrons forming the burst, which is caused by the Gaussian character of the distribution, has practically wiped out the rise in the spectrum of the elementary pulse due to a single electron (which is caused by the particular electrode structure).

Figure 5 has as abscissa the transit time relative to that for zero emission velocity. The latter is inversely proportional to the square root of the voltage. Furthermore, Malter's data⁴ indicate that the relative spread of emission velocities decreases with increasing primary velocity, so that raising the inter-stage voltages would decrease the spread in transit-time somewhat more than in proportion to the square root of the voltage. Unlike the frequency dependence due to the finite velocities of the electrons, which depends only on conditions in the last two stages, the frequency dependence due to spread in transit time depends on conditions in all the

stages from (1) on. Both effects can be pushed up to higher frequencies by increasing the accelerating voltages.

By the theorem in Fourier analysis already referred to, the spectrum of the noise is the product of the two spectra we have just calculated. It is plotted in Fig. 7. At 240 Mc, the output is down 1 db. It reaches 6 db down at 570 Mc, and falls more and more rapidly as the frequency increases. The spread in time of arrival has wiped out, for all practical purposes, the resonance effect due to the two-sidedness of the elementary pulse; the effect of the shape of the spectrum of the elementary pulse is just barely perceptible.

It is clear that no great quantitative significance can be claimed for our final curve, because of the roughness of the data and the method of calculation, as well as the variations in behavior from tube to tube. It does seem safe to state however, that the noise spectrum should be flat up to about 100 Mc, begin to fall off perceptibly in the range 100 to 1000 Mc, and be very weak at higher frequencies. The null at around 3300 Mc would occur in an output which is already very weak.

Some confirmation of these predictions is provided by unpublished measurements of P. Jastram at the Radio Research Laboratory. They showed that the noise output around 30 Mc is about the same as that in the video range. There is also some experimental evidence to indicate that at 100 Mc the output is still about the same, and that above this frequency, it becomes much smaller.

I am grateful to Doctors J. H. Van Vleck and S. A. Goudsmit for illuminating discussions.

The Response of Biased, Saturated Linear and Quadratic Rectifiers to Random Noise*

DAVID MIDDLETON**

Radio Research Laboratory, Harvard University, Cambridge 38, Massachusetts

(Received May 3, 1946)

Spectral distributions and powers associated with the output of biased, saturated linear and quadratic rectifiers are determined according to the method of Rice when the incoming disturbance is random noise. Three classes of input spectra are considered: (I), *broad band noise*, where the central frequency is equal to or less than the spectral width, (II), *semi-broad band noise*, for which the central frequency exceeds the width by a moderate amount, and (III), *narrow band noise*, where the width is much less than the central frequency. Rectification of types I and II yields spectra having roughly the same distribution as that of the incoming waves, but for type III an infinite number of separate noise bands are generated and appear in the output, centered about harmonics of the central frequency. It is found that clipping, whether at the "top" or "bottom" of the incoming wave, always spreads the spectrum and

reduces the output power. Further, it is shown that clipping due to cut-off alone produces a greater spectral spread than clipping with saturation in addition, for types I and II, but not necessarily for spectra of type III. Symmetrical clipping for classes I and II yields little broadening of the spectrum, even in extreme cases, and for class III waves, the even-harmonic regions are completely missing except for a d.c. component, and only the odd-harmonic zones appear. The behavior of linear and quadratic rectifiers is qualitatively similar in most cases. *The powers in the d.c. and continuous portions of the output spectrum are shown to be independent of the spectral shape of the incoming noise.* Some examples of extreme clipping are considered, numerous curves are included, and a general analysis is outlined.

I. INTRODUCTION

SINCE the nature of noise problems is basically statistical, one is never able to predict precisely the amplitude of the input or output noise wave at any given instant. Quantitatively the most that can be done in this respect is to determine the probability of occurrence of noise peaks, verify that they follow some law of distribution, and compute the values of the average, the mean-square, and the higher moments. A more far-reaching consequence of the statistical character of the noise lies in the fact that there exists no relationship between the noise power (or mean-square-amplitude) spectrum and the law of the distribution of amplitudes by which one may be obtained from the other. For example, the spectra of clipped† and unclipped noise may be identical, while the instantaneous voltage waves whose power spectra we are comparing are quite different, as may be verified experimentally by observation on a cathode-ray

oscilloscope. The former disturbance will show a definite ceiling where clipping has occurred, while the latter will be jagged and without a definite boundary. There is no bridge between the average and the instantaneous in problems of this type because information about the phases of the components is always lacking when we seek to pass from the spectral to the instan-

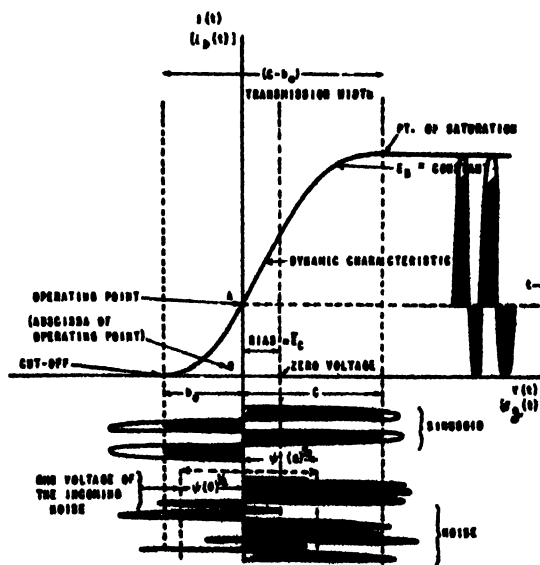


FIG. 1. General dynamic characteristic for resistive loads.

* This paper is based on work done for the Office of Scientific Research and Development under Contract No. OEMar-411 with the President and Fellows of Harvard College.

** Now located at the Research Laboratory of Physics, Harvard University, Cambridge 38, Massachusetts.

† By clipping we mean amplitude-limiting, either at the top or bottom of the wave or both.

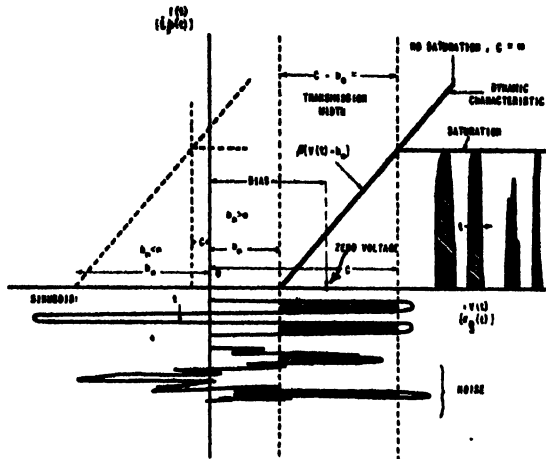


FIG. 2a. Dynamic characteristic for the biased and saturated linear rectifier.

taneous amplitude distribution. Thus for a given probability distribution of amplitudes any number of differently shaped spectra are possible, and *vice versa*.

Now when random noise enters a non-linear device the output disturbance is distorted, and a change in the amplitude distribution and spectrum of the incoming wave results. Cross-modulation, or mixing, of the various components occurs, so that the outgoing wave has generally quite different instantaneous and average characters from those of the incoming signal. For example, the statistics of the amplitude distribution are modified, the noise is no longer "pure" or random in the above sense, and the power and the power spectrum of the output in many instances differ noticeably from those of the input. The factors determining the character of the outgoing wave are chiefly (1), the rectifier response, which is the dynamic $i_p - e_0$ characteristic of the tube in question, (2), the spectrum of the incoming random noise, and (3), the cut-off and saturation levels, which specify the extent of the clipping, as illustrated for the general case in Fig. 1, and specifically for the linear and a general form of quadratic rectifier in Figs. 2a and 2b.

The form of the dynamic response, and to a less extent the output current at saturation, depend also on the load in the plate circuit of the rectifier. The analysis as presented here applies for purely resistive loads, and is a good approximation for more general impedances,

provided the reactance X is small compared with the resistance R . For appreciable reactances, however, the analysis becomes much more intricate.

Our rectifier is operated at some bias such that the cut-off is b_0 volts and saturation occurs at C volts, both measured from the abscissa of the operating point 0, as indicated in Fig. 1, where C always exceeds b_0 . The (voltage) *transmission width* of the device is defined as that region of the response curve which will pass a portion of the incoming wave. Analytically, as can be observed from Figs. 1 and 2, the transmission width is $C - b_0$ for the linear detector, and $2(C - b_0)$ for the quadratic response. For our "linear" device the output current, or voltage measured across an impedance, is then

$$\begin{aligned} I(t) &= \beta(C - b_0), & C \leq V, \\ &= \beta(V - b_0), & V \leq C, \\ &= 0, & V \leq b_0, \end{aligned} \quad (1-1)$$

vide Fig. 2a, where the quantity β is the (dynamic) tube transconductance. When the curvature of the $I - V$ characteristic is small in the vicinity of cut-off and saturation, the linear dynamic path is a fair approximation for a large class of actual operating conditions.

We distinguish two types of quadratic response, which we designate as Type A and Type B. The Type A characteristic, similar to the Type B except that it levels off abruptly at $V = C$, Fig. 2b, is less important and less general than the Type B, and so will not be considered further in this paper. The Type B characteristic, obtained analytically by fitting two parabolas together in the manner indicated in Fig. 2b, is

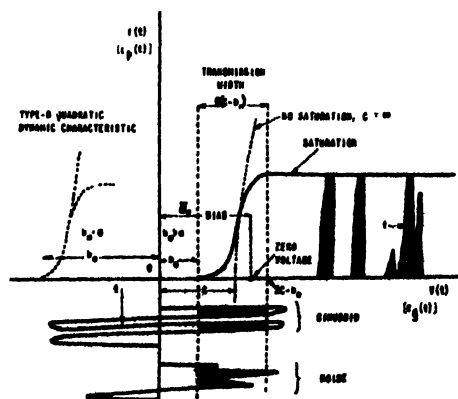


FIG. 2b. Dynamic characteristic for the biased and saturated Type-B quadratic rectifier.

perhaps the simplest theoretical representation of a general form of dynamic response often encountered in practice, particularly in that curvature in the neighborhood of cut-off and saturation is considered. Then for Type B tubes we may write the output current as

$$\begin{aligned} I(t) &= 2\lambda^2(C-b_0)^2, \quad 2C-b_0 \leq V, \\ &= \lambda^2\{2(C-b_0)^2 - (V-2C+b_0)^2\}, \\ &\quad C \leq V \leq 2C-b_0 \quad (1-2) \\ &= \lambda^2(V-b_0)^2, \quad b_0 \leq V \leq C, \\ &= 0, \quad V \leq b_0, \end{aligned}$$

where now λ is a scale factor with the dimensions of mhos.

The purpose of this paper is to investigate both qualitatively and quantitatively the effects of cut-off and saturation in determining the output spectra and powers when random noise enters a non-linear device similar to that considered above, shown in Fig. 2. The theory is developed following a method originally suggested by S. O. Rice of the Bell Telephone Laboratories.¹ The assumption is made that *initially* the noise is random, with a normal or Gaussian distribution of amplitudes. In practice this is never precisely true, as the larger noise peaks are inevitably limited to some extent by nominally linear circuit elements such as amplifiers. However, when the limiting occurs but a small percentage of the time, our "pure" noise may then be substituted for the actual disturbance with little error and the results so obtained apply in the practical case.

II. REMARKS ON THE METHOD OF ANALYSIS

To calculate the power and spectra for the noise¹⁻³ after it has passed through a non-linear device we use the contour representation^{1,3-5} of

the output current $I(t)$ as a function of the input voltage $V(t)$, namely

$$I(t) = \frac{1}{2\pi} \int_C f(iz) e^{i\psi(t)z} dz, \quad (2-1)$$

where C is a contour extending from $-\infty$ to $+\infty$ along the real axis and is indented downward about a possible singularity at the origin. The function $f(iz)$ is the Fourier transform of the rectifier characteristic and usually has a simple pole or branch-point at $z=0$.

Now if the correlation function $R(t)$ of the rectifier output is known we may calculate the power spectrum $W(f)$ from the important relation, deduced originally by Wiener,⁷ and later by Khintchine,⁸ and others,⁹ viz:

$$W(f) = 4 \int_0^\infty R(t) \cos \omega t dt, \quad \omega = 2\pi f, \quad (2-2)$$

which has the transform

$$R(t) = \int_0^\infty W(f) \cos \omega t df. \quad (2-3)$$

It is also noted that the mean total power is simply $R(0)$. For an input disturbance consisting solely of noise the correlation function for the output takes the form¹⁰

$$\begin{aligned} R(t) &= \frac{1}{4\pi^2} \int_C f(iz) \exp(-\psi z^2/2) dz \\ &\times \int_C f(i\xi) \exp(-\psi \xi^2/2 - \psi(t)z\xi) d\xi, \quad (2-4) \end{aligned}$$

where $\psi = \psi(0)$ is the mean square input voltage and $\psi(t)$ is the input correlation, having associated with it, in virtue of (2-2), the input spectrum $w(f)$.

III. OUTPUT POWER AND SPECTRA

If, after substituting the value of $R(t)$ from (2-4) in (2-2), we expand $e^{-\psi(t)z\xi}$ in powers of z and ξ , we may write the spectrum $W(f)$ of the

¹ S. O. Rice, "Mathematical analysis of random noise," Bell Sys. Tech. J. 23, 282 (1944); 24, 46 (1945).

² M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, No. 2 and 3, 323 (1943).

³ D. O. North, "The Modification of Noise by Certain Non Linear Devices," a synopsis of which was presented at the Jan. 1944 Winter Technical Meeting of the I.R.E.

⁴ W. R. Bennett and S. O. Rice, "Note on methods of computing modulation products," Phil. Mag. 18, 422 (1934).

⁵ W. R. Bennett, "Response of a linear rectifier to signal and noise," J. Acous. Soc. Am. 15, 164 (1944). (See also Bell Sys. Tech. J. 23, 97 (1944).)

⁶ K. Fränzl, Zeits. f. Hoch. u. Eleck. 57, 146 (1941), and Eleck. Nach. 17, 215 (1940); 19, 285 (1942).

⁷ N. Wiener, Acta. Math. 55, 117 (1930).

⁸ Khintchine, Math. Ann. 109, 604 (1934).

⁹ See footnote 12 of reference 3, and also reference 1 Part II.

¹⁰ This is a special case of Eq. (4.8-7) reference 1, in which $g(u, v, \tau)$ is set equal to unity.

TABLE I. Comparison of the precise and series* values of the total power output, for the linear rectifier.

$b_0/\psi^{\frac{1}{2}}$	$C/\psi^{\frac{1}{2}}=0.5$			-1.0			-1.5			-2.0			∞		
	7 Terms	14 Terms	$R(0)$	7 Terms	14 Terms	$R(0)$	7 Terms	14 Terms	$R(0)$	7 Terms	14 Terms	$R(0)$	7 Terms	14 Terms	$R(0)$
-2.0	3.794	3.795	3.795	4.419	4.419	4.419	4.765	4.766	4.766	4.919	4.920	4.921	4.994	4.994	4.994
-1.5	3.232	3.235	3.235	2.731	2.732	2.735	3.026	3.028	3.028	3.161	3.162	3.163	3.238	3.238	3.238
-1.0	1.108	1.112	1.123	1.505	1.507	1.516	1.751	1.753	1.755	1.887	1.890	1.890	1.843	1.924	1.925
-0.5	.417	.422	.435	.701	.706	.718	.895	.898	.900	.991	.996	.996	.943	1.055	1.05(6)
0.0	.0867	.0899	.0926	.251	.255	.258	.385	.388	.390	.458	.459	.460	.497	.499	.500
0.4	.0029	.0031	.0032	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.0	0.0	0.0	.0468	.0490	.0510	.122	.127	.128	.174	.178	.178	.207	.209	.210
0.9	—	—	—	.0014	.0015	.0017	—	—	—	—	—	—	—	—	—
1.0	—	—	—	0.0	0.0	0.0	.0301	.0217	.0232	.0488	.0517	.0526	.0733	.0746	.0753
1.4	—	—	—	—	—	—	.00055	.00065	.0007	—	—	—	—	—	—
1.5	—	—	—	—	—	—	0.0	0.0	0.0	.0073	.0080	.0086	.0220	.0224	.0228
1.9	—	—	—	—	—	—	—	—	—	.00018	.0002	.0002	—	—	—
2.0	—	—	—	—	—	—	—	—	—	0.0	0.0	0.0	.0053	.0056	.0058

$R(0)$ is in units of $\sigma^2\psi(0)$ and is obtained from the exact, closed form.

* The series value is obtained by evaluation of $R(0) \doteq \sum_{n=0}^N \frac{\psi(0)^n}{n!} h_{0,n}^2 (-1)^n$ where N is theoretically infinite, but in practice for these computations is 14. Calculations for $N=7$ are also given above.

output in a form convenient for calculation:

$$W(f) = \sum_{n=0}^{\infty} (-1)^n c_{0,n}(f) h_{0,n}^2 / n!, \quad (3-1)$$

where

$$c_{0,n}(f) = 4 \int_0^{\infty} \psi(t)^n \cos \omega t dt, \quad (3-2)$$

and

$$h_{0,n} = \frac{1}{2\pi} \int_C z^n f(iz) \exp[-\psi(0)z^2/2] dz. \quad (3-3)$$

Note that when $n=0$ we obtain the d.c. power. Now from (2-2) and (3-1)–(3-3) the correlation of the output is

$$R(t) = \sum_{n=0}^{\infty} (-1)^n h_{0,n}^2 \psi(t)^n / n!, \quad (3-4)$$

the d.c. power becomes

$$W_{d.c.} = h_{0,0}^2, \quad (3-5)$$

while the power in the continuum of the transmitted wave may be expressed by

$$W_c = \sum_{n=1}^{\infty} (-1)^n \psi(0)^n h_{0,n}^2 / n!. \quad (3-6)$$

We see from (3-6) that the power in the continuum, and for that matter in the d.c. component of the output disturbance, is independent of the spectral distribution of the input wave, and for a given current-voltage characteristic depends only on the mean-square input-noise voltage. This

result is a somewhat special case of a more general theorem; (see Eq. (4.9-19) of reference 1).

Unfortunately, the series (3-6) does not converge rapidly except when the output spectrum consists of separate spectral regions, a condition which arises if the input wave is narrow-band. In some cases we may use other methods of obtaining the mean power in the continuum. For the total mean output power we have

$$R(0) = W_r = (2\pi\psi)^{-1}$$

$$\int_{-\infty}^{\infty} I(V)^2 \exp(-V^2/2\psi) dV, \quad (3-7)$$

with $I(V)$, the output current in a unit resistance, given by (1-1) or (1-2). If we normalize V by the substitution $V = x\psi^{\frac{1}{2}}$, Eq. (3-7) may be written in the more useful form

$$R(0) = (2\pi)^{-1} \int_{-\infty}^{\infty} I(x\psi^{\frac{1}{2}})^2 \exp(-x^2/2) dx. \quad (3-8)$$

Equation (3-8) yields the total output power, and the power associated with the continuum may be found from (3-5) by subtraction, since

$$W_c = W_r - W_{d.c.} = R(0) - h_{0,0}^2. \quad (3-9)$$

Note that we may also obtain the d.c. power from

$$W_{d.c.} = \left\{ (2\pi)^{-1} \int_{-\infty}^{\infty} I(x\psi^{\frac{1}{2}}) \exp(-x^2/2) dx \right\}^2. \quad (3-10)$$

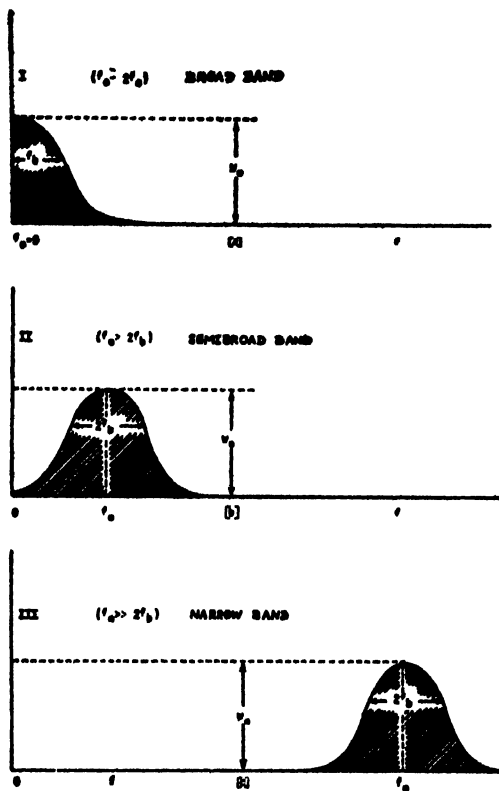


FIG. 3. The three classes of input spectra.

Table I illustrates the numerical difference between the series computation which uses (3-6) and the "precise" value (3-8).

It is convenient to classify input spectra roughly in three groups: (I) *broad band noise*, which is characterized by a band width as large as or greater than the central frequency f_0 of its spectrum, (II) *semi-broad band noise*, where the central frequency is of the same order of magnitude as the spectral width, but always larger, and finally (III), *narrow-band noise*, for which the central frequency is decidedly greater than the band width. Figure 3 illustrates I-III. It is clear that groups I-III may overlap; the extent to which overlapping occurs depends on our choice of what constitutes "band width." To facilitate the calculations in some cases, and to preserve a reasonably clear-cut distinction between classes I-III, we set as our spectral limits determining the band width the frequencies at which the (power) spectral level falls to a few percent of its maximum value. Our classification depends of course, on the assumption that the

input spectra have but one maximum, a condition quite common in practice. The analysis is perfectly capable of handling more involved spectral shapes, but the computational difficulties are then greatly increased.

The input spectra considered specifically in this paper have the general form

$$w(f) = W_0 \exp [-(\omega - \omega_0)^2 / \omega_b^2], \quad \omega \geq 0, \quad (3-11)$$

where $\omega_0 = 2\pi f_0$ is the angular central frequency, ω_b is an angular frequency relating the width to the central frequency, and hence governing the relative spread of the input spectrum, and W_0 is the maximum spectral density, in units of volts²/frequency. Our choice of (3-11) is based on the observation that the Gaussian shape is a simple and reasonably good approximation of many spectra encountered in practice, particularly for disturbances passing through single or cascaded tuned circuits, in receivers, transmitters, and similar apparatus. Actually, as Wallman¹¹ has shown, there are no physical filters (i.e., frequency-discriminating devices) that can be constructed with a finite number of passive elements to give precisely a Gaussian response, for only with an infinite time delay is it possible to achieve the rapid falling-off of the response that distinguishes (3-11) for frequencies moderately well away from f_0 , for, let us say, $f_0/f_b > 2$. Spectra of the type (3-11), however, appear to be closer approximations in practice than the so-called "ideal" or uniform band.

In what follows we shall find it more convenient to use the normalized input correlation function $r(t)$, rather than $\psi(t)$. Here we have

$$r(t) \equiv \psi(t) / \psi(0) \\ = \int_0^\infty w(f) \cos \omega t df / \int_0^\infty w(f) df; \quad (3-12)$$

then $r(0)$ is unity and $r(t) < 1$, $t > 0$. The input correlation for class I spectra ($f_0 = 0$), is easily deduced from (3-11) and (3-12). We have

$$r(t)_I = \exp (-\omega_b^2 t^2 / 4); \quad \psi(0) = W_0 \omega_b / 4\pi. \quad (3-13)$$

When $f_0 > 0$, $r(t)$ becomes

¹¹ H. Wallman, "Realizability of filters," Radiation Laboratory Report 637, M. I. T., Dec. 8, 1944.

$$r(t)_{II,III} = (\omega_b \pi^{\frac{1}{2}})^{-1} \int_0^{\infty} \exp [-(\omega - \omega_0)^2 / \omega_b^2] \cos \omega t df \\ = (2\pi^{\frac{1}{2}} \omega_b)^{-1} \int_{-\infty}^{\infty} \exp (-\omega'^2 / \omega_b^2) \cos (\omega' + \omega_0) t d\omega', \quad (3-14)$$

where now

$$\psi(0) = W_0 \omega_b / 2\pi^{\frac{1}{2}}.$$

If there is an ignorable "tail" at $f=0$, we may extend the lower limit of integration to $-\infty$ with negligible error and obtain

$$r(t)_{II,III} = \cos \omega_0 t \exp (-\omega_b^2 t^2 / 4) \\ = r_0(t) \cos \omega_0 t, \quad (3-15)$$

which defines $r_0(t)$, a useful quantity in later work. Note from (3-14) that $\psi_{II,III} = 2\psi(0)_I$ since both sides, $f \gtrless f_0$, contribute to the mean power. We shall consider here only examples of class II spectra for which (3-15) applies.

When the incoming wave is a narrow band and has the spectrum (3-11), Eq. (3-14) may be used with even less error than occurs for the semi-broad band case, since here f_0 is very much larger than f_b .

To determine the output spectrum for (1), we see from (3-2) and (3-13) that

$$c_{0,n}(f)_I = 4\psi^n (\pi/n)^{\frac{1}{2}} \frac{\exp (-\omega^2 / n\omega_b^2)}{\omega_b}. \quad (3-16)$$

Similarly from (3-2) and (3-15) the semi-broad band case yields

$$c_{0,n}(f)_{II} = 4\psi^n \int_0^{\infty} \exp (-n\omega_b^2 t^2 / 4) \\ \times (\cos \omega_0 t)^n \cos \omega t dt, \quad (3-17)$$

and this may be evaluated with the help of

$$(\cos \omega_0 t)^n = 2^{1-n} \sum_{j=0}^{n/2, (n-1)/2} {}_n C_j {}_n \Delta_j \cos (n-2j)\omega_0 t \quad (3-18)$$

where

$${}_n \Delta_{n/2} = \frac{1}{2}, \quad {}_n \Delta_j = 1; \quad n \neq 2j, \quad \text{and} \quad {}_n C_j = n! / (n-j)! j!$$

is the familiar binomial coefficient. The limits on the summation apply for even or odd values of n , respectively. Substitution of (3-18) in (3-17) leads to integrals of the type

$$\int_0^{\infty} \exp (-n\omega_b^2 t^2 / 4) \{ \cos [(n-2j)\omega_0 + \omega] t \\ + \cos [(n-2j)\omega_0 - \omega] t \} dt,$$

which appears to be resolved into components representing separate spectral zones distributed about $(n-2j)f_0$. However, since f_b and f_0 are comparable in magnitude, one "harmonic" zone $l=l'$, corresponding to all values of $(n-2j)$ which are equal to l , overlaps another zone, $l=l''$, etc., so that physically the output spectrum cannot be resolved into distinct harmonic noise bands. It is represented by the superposition of all the different zones $l \geq 0$. We may apply (3-18) nevertheless, to obtain finally

$$c_{0,n}(f)_{II} = 4\psi^n (\pi/n)^{\frac{1}{2}} \frac{1}{\omega_b} K_n(\omega/\omega_b) \\ \times \exp (-\omega^2 / n\omega_b^2), \quad (3-19)$$

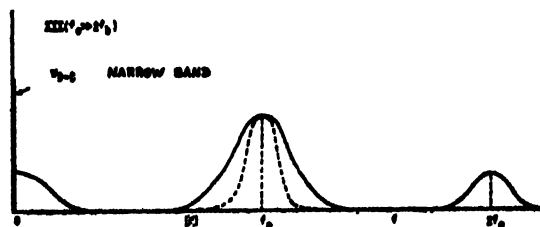
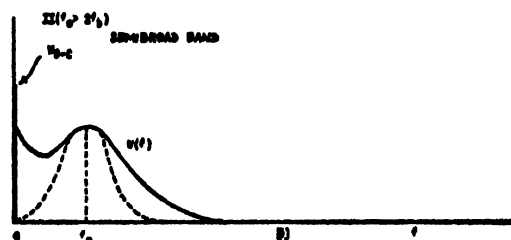
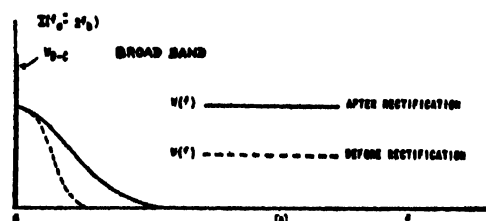


FIG. 4. The three classes of input spectra after rectification, arbitrary normalization.

where

$$K_n(\omega/\omega_b) = 2^{1-n} \sum_{j=0}^{n/2, (n-1)/2} n \Delta_j C_j \times \cosh \left[\frac{2(n-2j)\omega_0\omega}{n\omega_b^2} \right] \times \exp \left[-(n-2j)^2\omega_0^2/n\omega_b^2 \right]. \quad (3-20)$$

Spectral "tails" at $f=0$ are ignored, as they are relatively insignificant.

When narrow band noise (III) enters our non-linear device, a separation of the harmonic zones becomes physically possible, since now $f_0 \gg f_b$. By the substitution of Eqs. (3-18) in (3-17) we are able to pick out the various spectral regions analytically; in practice we use filters having suitable response and centered about lf_0 .* For a given value of l we require all n for which $j=(n-l)/2 \geq 0$, where n is odd for odd l , and even for even l . Then (3-17) may be modified to give

$$c_{0,n}^l(f)_{III} = \left[4\psi^n n! \Delta_{(n-l)/2} / \left(\frac{n-l}{2} \right)! \left(\frac{n+l}{2} \right)! 2^{n-1} \right] \int_0^\infty r_0(t)^n \cos \omega t \cos l\omega_0 t dt \\ - \left[4\psi^n n! \Delta_{(n-l)/2} / \left(\frac{n-l}{2} \right)! \left(\frac{n+l}{2} \right)! 2^{n-1} \right] \times \{ \exp [-(l\omega_0 + \omega)^2/n\omega_b^2] + \exp [-(l\omega_0 - \omega)^2/n\omega_b^2] \} / 2. \quad (3-21)$$

Because $lf_0 \gg f_b$, we may safely ignore the term $\exp [-(l\omega_0 + \omega)^2/n\omega_b^2]$, except when $l=0$. Figure 4 illustrates the form of the spectral output for classes I-III. The normalization of the spectra is arbitrary.

By taking $l=n=0$ we obtain the d.c. output power,

$$W_{d.c.} = h_0 \sigma^2, \quad \text{since} \quad \int_0^\infty c_{0,0}^0(f) df = 2 \int_0^\infty \delta(f-0) df = 1. \quad (3-22)$$

The correlation function corresponding to the l th zone is

$$R_l(t) = \epsilon_l \cos l\omega_0 t \sum_{n=l, l+2, \dots}^{\infty} (-1)^n \psi^n h_0 \sigma^2 r_0^n(t) / \left(\frac{n-l}{2} \right)! \left(\frac{n+l}{2} \right)! 2^n, \quad (3-23)$$

where

$$\epsilon_l = 2, \quad l \geq 1; \quad \epsilon_0 = 1, \quad \text{and} \quad r_0(t) = \exp(-\omega_b^2 t^2/4).$$

When we take $t=0$ it follows that the mean power in the low frequency continuum is determined provided $n=2, 4, \dots, l=0$:

$$W_{l,t.} = \sum_{n=l, l+2, \dots}^{\infty} \psi^n h_0 \sigma^2 / 2^n (n/2)!^2. \quad (3-24)$$

Equation (3-23) also enables us to find the power associated with each of the higher ($l \geq 1$) spectral regions. Setting $t=0$ in (3-23) gives

$$W_l = \sum_{n=l, l+2, \dots}^{\infty} (-1)^n \psi^n h_0 \sigma^2 / 2^{n-1} \left(\frac{n-l}{2} \right)! \left(\frac{n+l}{2} \right)!, \quad l \geq 1. \quad (3-25)$$

Note that W_l is independent of the spectral distribution of the input wave.¹³

* Note: here we tacitly assume "ideal" filters, with a uniform response in the regions of interest, so that all components of a particular harmonic band are passed without distortion. When this is not the case the output is $|W(f)|^2 |G(f)|^2$, where $G(f)$ is the amplitude response of the filter in question.

¹³ This is an example of Eq. (4.9-19), reference 1.

IV. THE BIASED, SATURATED LINEAR RECTIFIER

The explicit calculation of powers and spectra depends on a knowledge of $h_{0,n}$, Eq. (3-3). In particular we are interested in the general biased and saturated linear rectifier shown in Fig. 2, for which the Fourier transform, i.e., $f(is)$, of the $i_p - e_s$ characteristic (Eq. 1-1) is readily demonstrated to be

$$f(is) = \beta [\exp(-ib_0 z) - \exp(-iCz)] / (iz)^2, \quad C \geq b_0, \quad (4-1)$$

so that (3-3) becomes

$$h_{0,n} = -\frac{\beta}{2\pi} \int_C z^{n-2} [\exp(-ib_0 z) - \exp(-iCz)] \exp(-\psi z^2/2) dz, \quad C \geq b_0. \quad (4-2)$$

This integral is evaluated in the Appendix, and yields the following results:

$$h_{0,0} = -\beta\psi^{1/2} \{ (\bar{C}\Theta_c - 1) - \bar{b}(\Theta_{b_0} - 1) + 2(\phi_c - \phi_{b_0}) \}, \quad (4-3)$$

$$h_{0,1} = -\beta i / 2 \{ \Theta_c - \Theta_{b_0} \}, \quad (4-4)$$

$$h_{0,n} = \beta(-1)^{(n+2)/2} \psi^{(1-n)/2} \{ \phi_c^{(n-2)} - \phi_{b_0}^{(n-2)} \}, \quad n = 2, 4, 6, \dots, \quad (4-5)$$

$$h_{0,n} = \beta i(-1)^{(n+3)/2} \psi^{(1-n)/2} \{ \phi_c^{(n-2)} - \phi_{b_0}^{(n-2)} \}, \quad n = 3, 5, 7, \dots \quad (4-6)$$

In the above, and for subsequent use, we have adopted the abbreviations

$$\bar{b}_0 = b_0/\psi^{1/2}; \quad \bar{C} = C/\psi^{1/2}; \quad \Theta_{b_0} = \Theta(b_0/(2\psi^{1/2})); \quad \Theta_c = \Theta(C/(2\psi^{1/2})), \quad (4-7)$$

and

$$\phi_c^{(j)} = \phi^{(j)}(C/\psi^{1/2}); \quad \phi_{b_0}^{(j)} = \phi^{(j)}(b_0/\psi^{1/2}), \quad j = 0, 1, 2, \dots$$

where ϕ is the same as $\phi^{(0)}$. The quantity Θ denotes the familiar error function, here written as

$$\Theta(x_0) \equiv (2/\pi^{1/2}) \int_0^{x_0} \exp(-y^2) dy; \quad \Theta(x_0) = -\Theta(-x_0), \quad (4-8)$$

and $\phi^{(j)}$ takes the form

$$\phi^{(j)}(x_0) = \frac{d^{(j)}}{dx^{(j)}} \exp(-x^2/2) \Big|_{x=x_0} = (-1)^j H_j(x_0) \exp(-x_0^2/2) / (2\pi)^{1/2}, \quad (4-9)$$

$H_j(x_0)$ being an Hermitian polynomial of the j th order. Extensive curves illustrating the effects of cut-off and saturation on the output spectra after rectification by a biased and saturated linear detector are given in Figs. 5-17,* based on (3-16), (3-19), and (3-21) in (3-1), when $h_{0,n}$ is obtained from (4-3)-(4-6).

In many cases the power output is of interest. Because of the relatively slow convergence of the series (3-4) and (3-6) we use instead the result (3-8) to compute the mean total output power W_r . For our device, characterized by Eq. (1-1), (or (4-1)), we see that

$$W_r = R(0) = \frac{\beta^2 \psi}{(2\pi)^{1/2}} \left\{ \int_{b_0/\psi^{1/2}}^{\bar{C}} (x - \bar{b}_0)^2 \exp(-x^2/2) dx + \int_{\bar{C}}^{\infty} (\bar{C} - \bar{b}_0)^2 \exp(-x^2/2) dx \right\}, \quad (4-10)$$

irrespective of whether or not the input spectrum belongs to groups I-III. Equation (4-10) integrates to

$$W_r = \beta^2 \psi \{ (\bar{C} - \bar{b}_0)^2 / 2 + (1 + 2\bar{C}\bar{b}_0 - \bar{C}^2) \Theta_c / 2 - (1 + \bar{b}_0^2) \Theta_{b_0} / 2 + 2\bar{b}_0 \phi_c + \phi_c^{(1)} + \phi_{b_0}^{(1)} \}. \quad (4-11)$$

The d.c. output power follows from (3-5) and (4-3), and that in the continuum from (3-9) and the above relation (4-11). Figures 18-20 illustrate the dependence of W_r , $W_{d.c.}$, and W_c on $\bar{b}_0$ and $\bar{C}$, while Figs. 21 and 22 show W_l , calculated from (3-24) and (3-25), for $l=0$ and 1.

* Note: in the figures we have represented $W(f)$ formally by its equivalent $W(\omega/\omega_0)$.

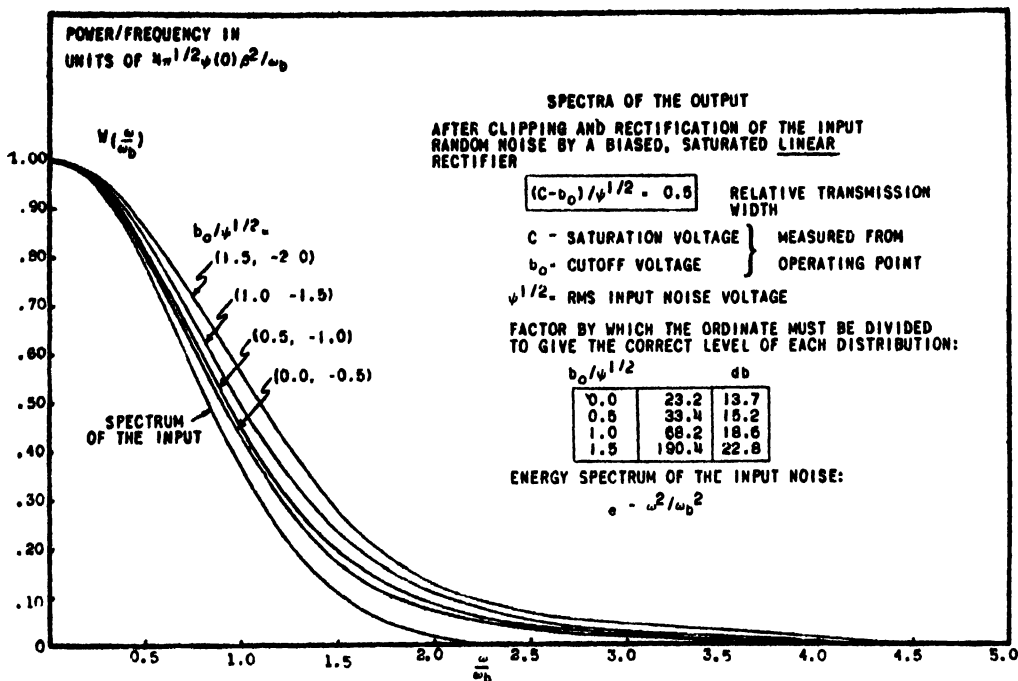


FIG. 5. Output spectra for broad band input noise (I) from a biased and saturated linear rectifier. (In this and succeeding figures, the correct output spectral level is expressed in units of $4\pi^{1/2}\psi(0)\beta^2/\omega_b$, after division by the factors indicated.)

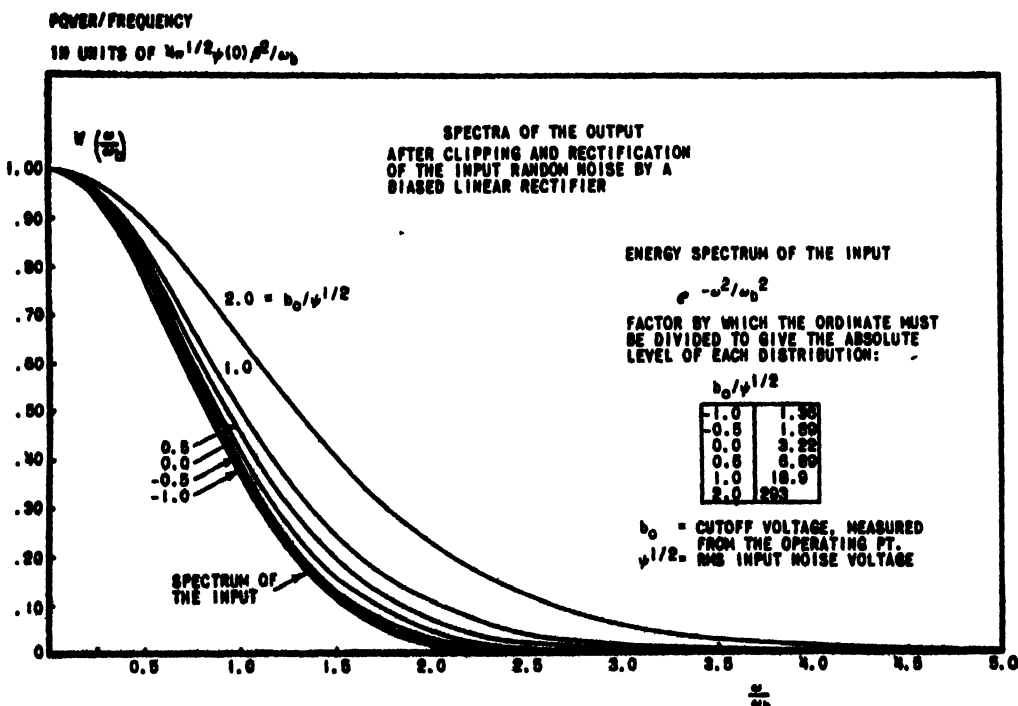


FIG. 6. Output spectra for broad band input noise (I) from a biased but unsaturated linear rectifier.

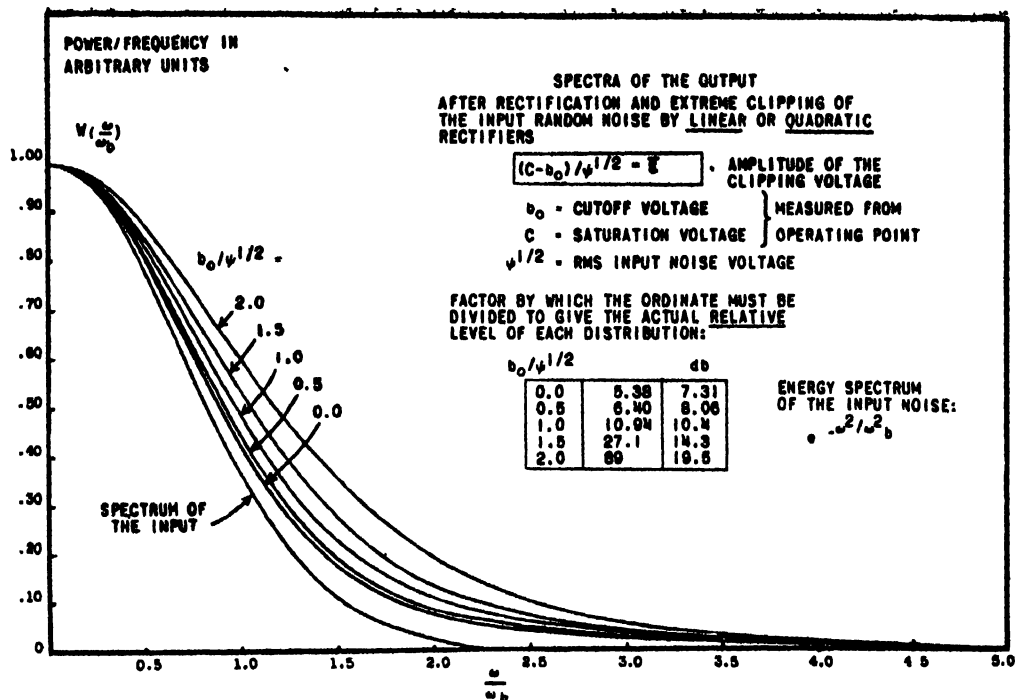


FIG. 7. Output spectra for broad band input noise (I) after extreme clipping in a linear rectifier. (The relative levels in this figure and in Figs. 9, 14, 17, are the levels expressed in the (same) arbitrary units, after division by the factors indicated.)

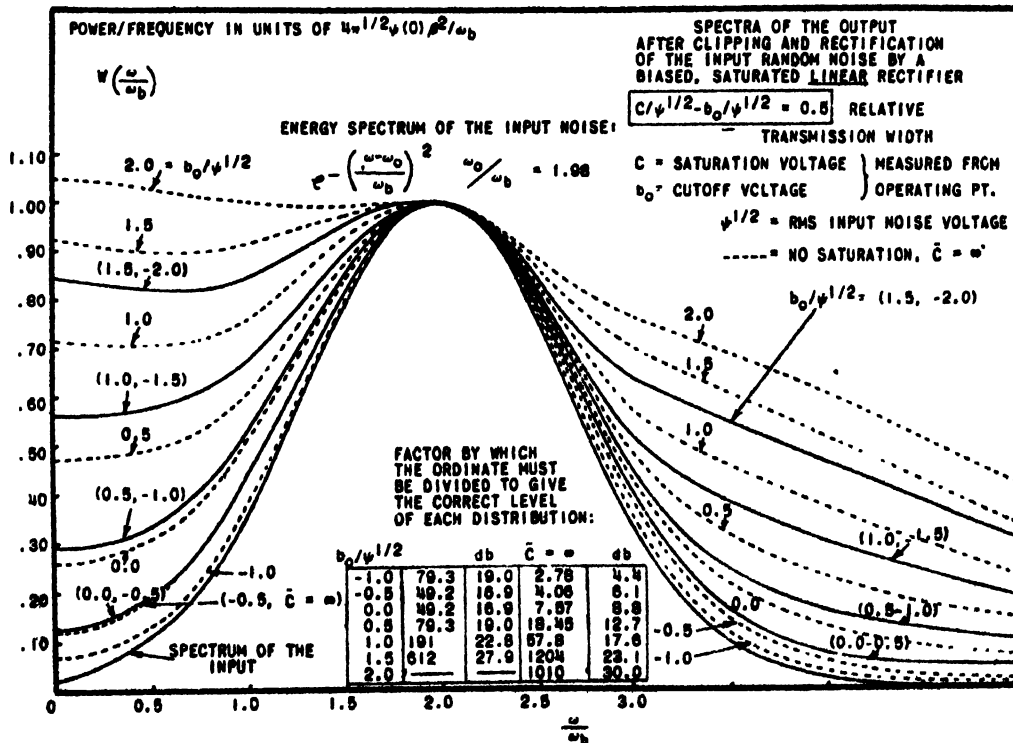


FIG. 8. Output spectra for semi-broad band noise (II), saturation and no saturation.

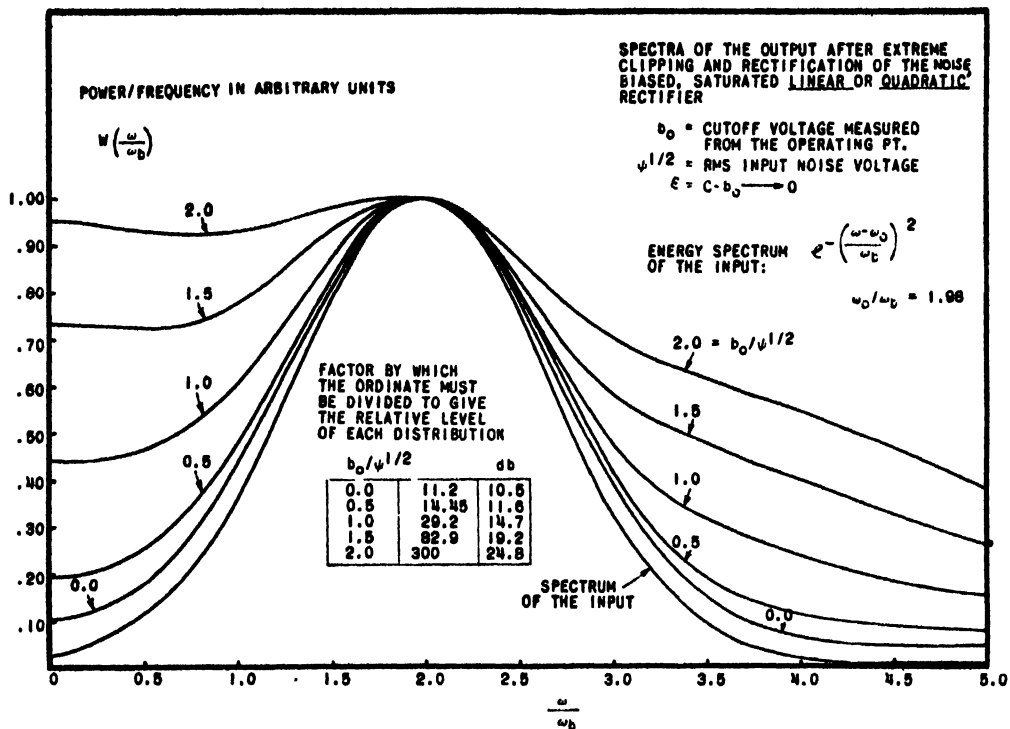


FIG. 9. Output spectra for semi-broad band noise (II) after extreme clipping.

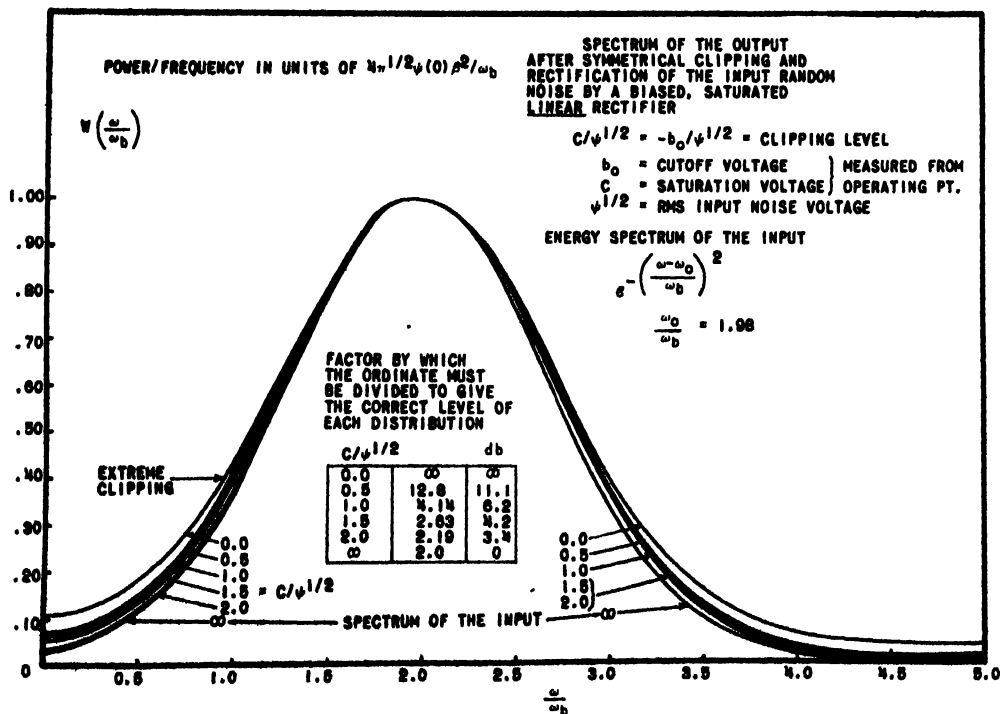


FIG. 10. Output spectra for semi-broad band noise (II) after symmetrical clipping.

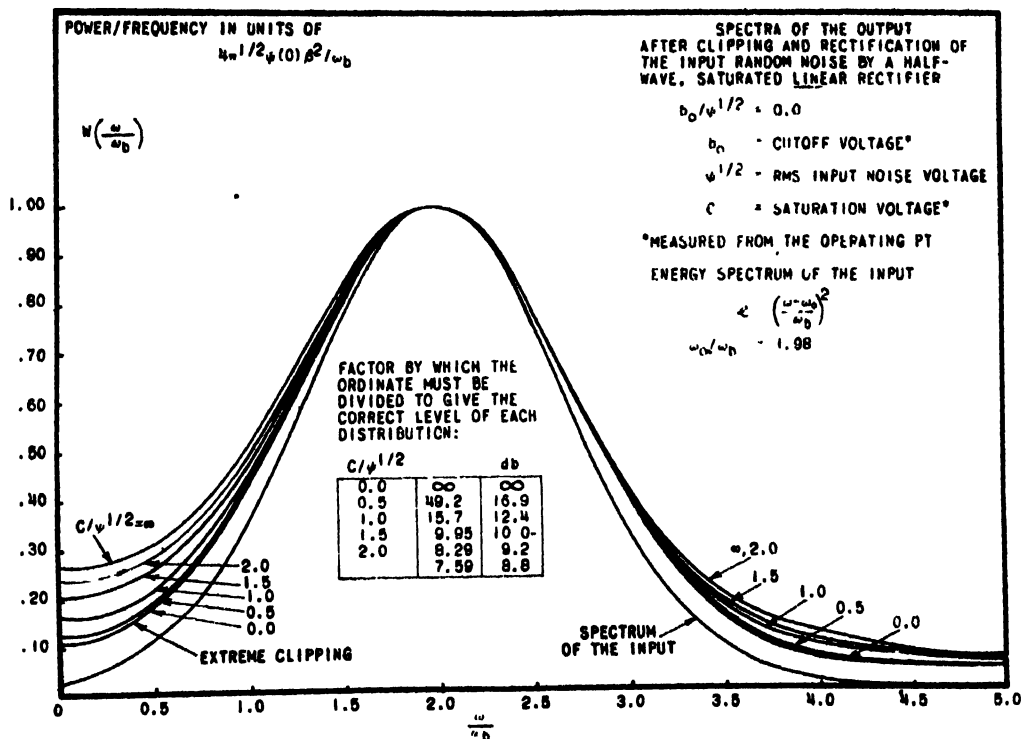


FIG. 11. Output spectra for semi-broad band noise (II) after half-wave rectification, with and without saturation.

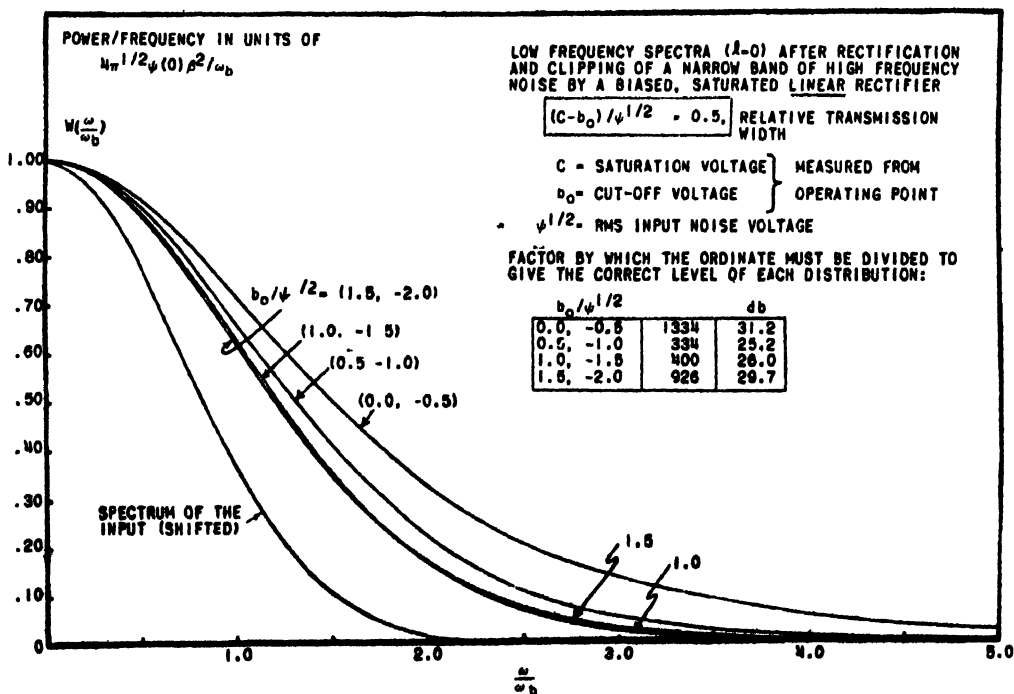


FIG. 12. Low frequency output spectra ($\lambda=0$), for narrow band input noise (III) with saturation.

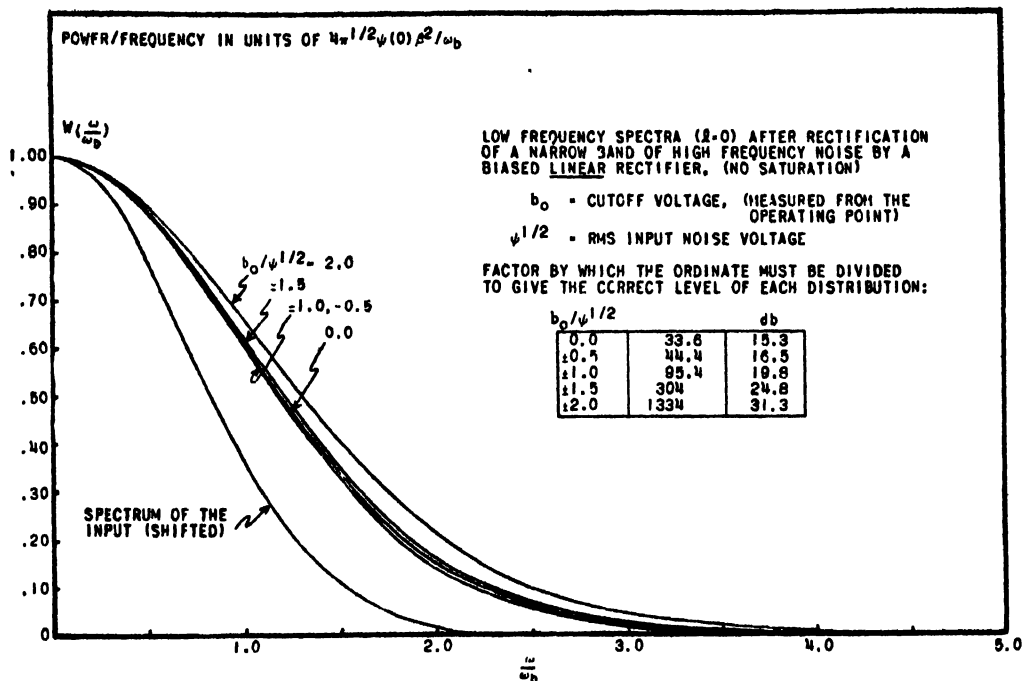


FIG. 13. Low frequency output spectra ($l=0$), for narrow band input noise (III) without saturation.

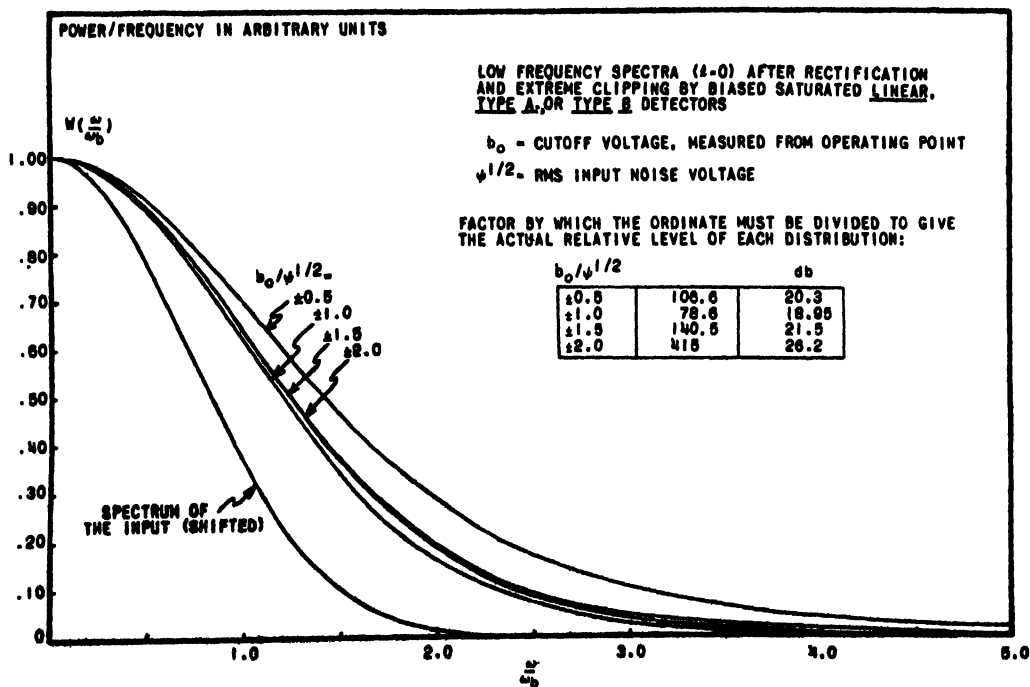


FIG. 14. Low frequency output spectra ($l=0$), for narrow band input noise (III) after extreme clipping.

FIG. 15. High frequency output spectra ($l=1$), for narrow band input noise (III) with saturation.

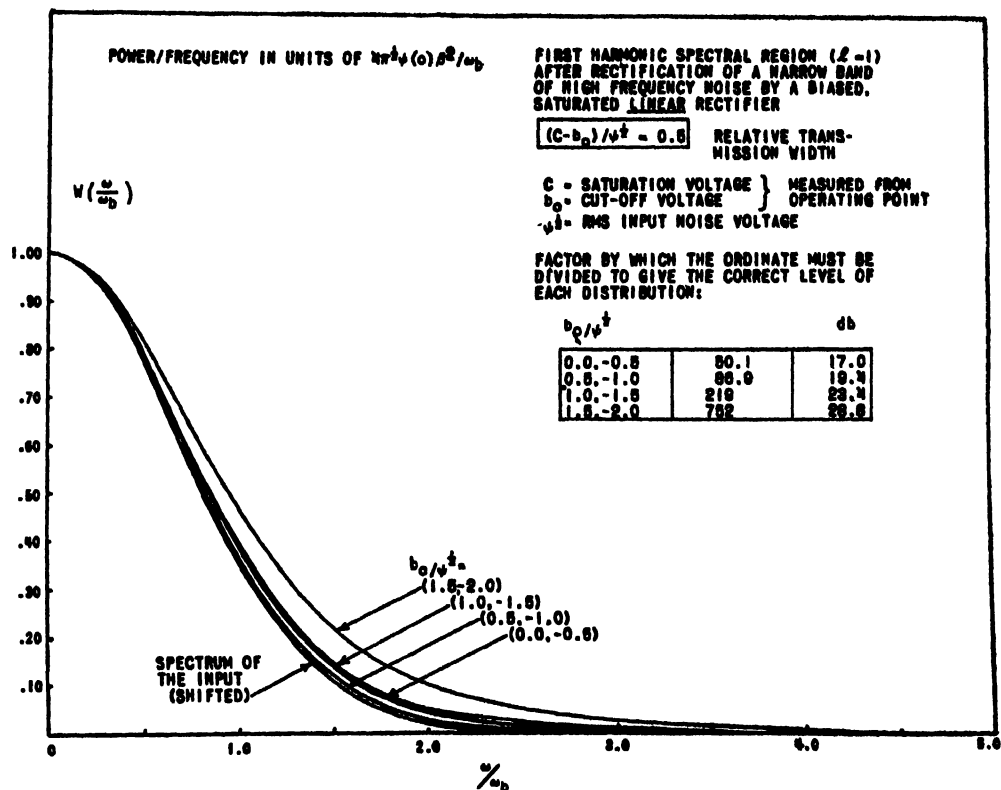
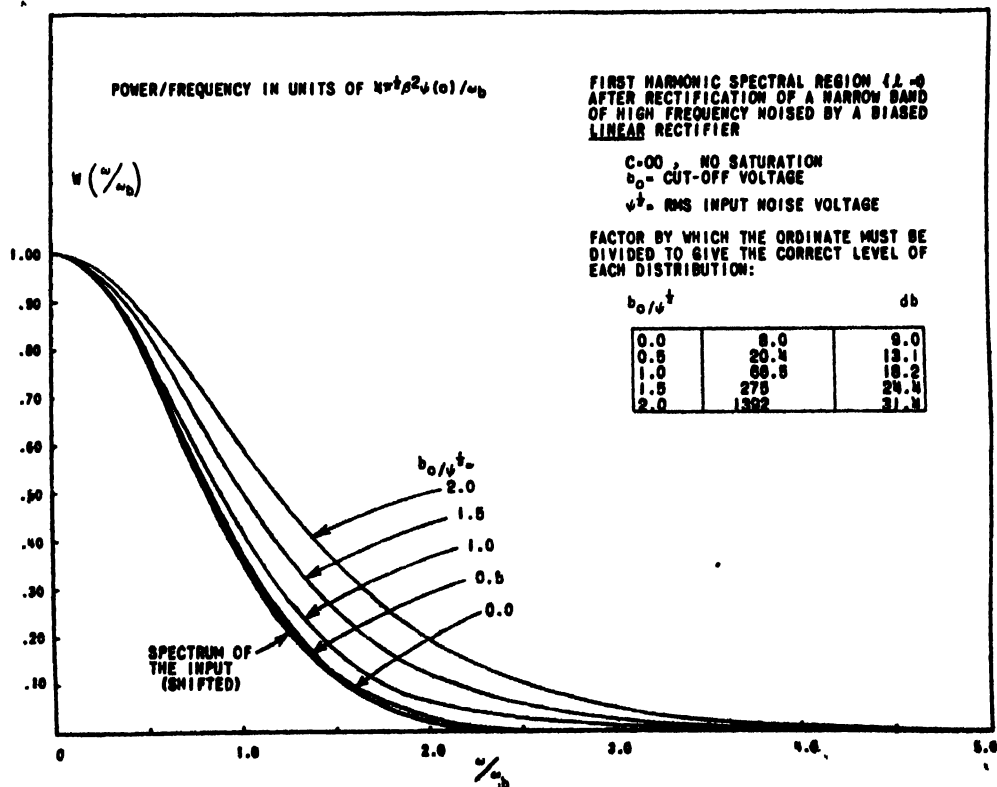


FIG. 16. High frequency output spectra ($l=1$), for narrow band input noise (III), no saturation.



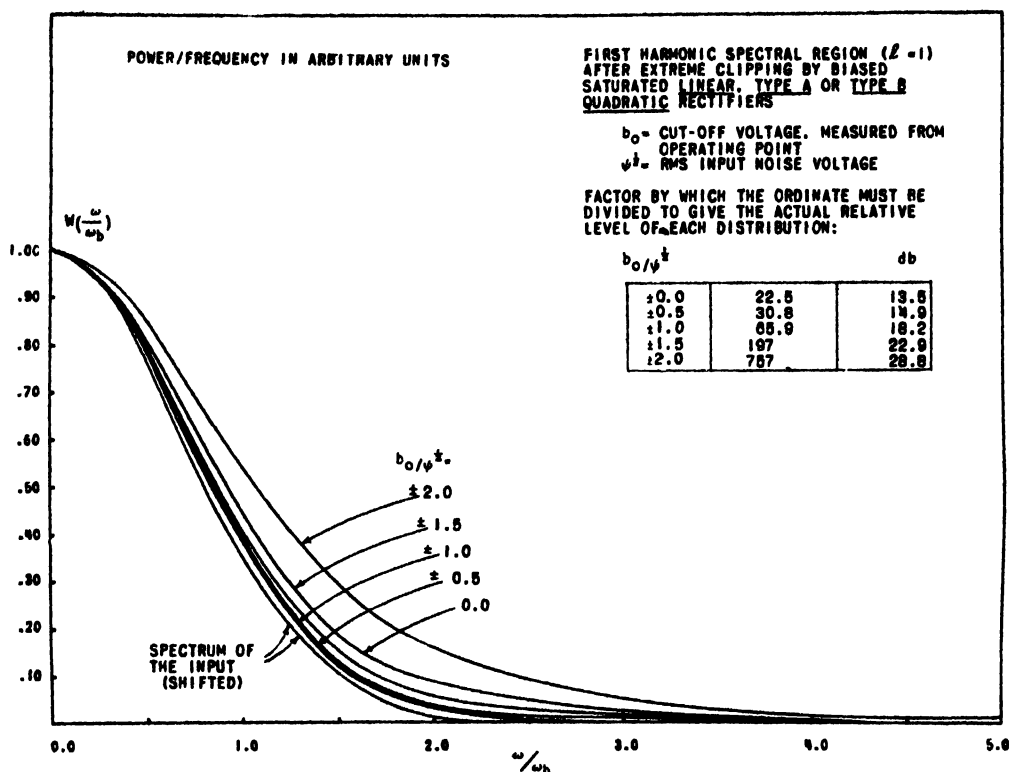


FIG. 17. High frequency output spectra ($l=1$), for narrow band input noise (III), extreme clipping.

A number of interesting special cases occur frequently in practice. The first is that of no saturation, a condition such that the instantaneous amplitudes exceed the critical value at which saturation occurs a negligible percentage of the time. Analytically $\bar{C}$ is then infinite, and Eqs. (4-3)–(4-6) are modified accordingly. Figures 6, 8, 13, 16 illustrate spectra in this instance; the dotted curves in Fig. 8 apply when $\bar{C} \rightarrow \infty$.

The second case is that of half-wave rectification where all input voltages V less than 0 are not passed, and all those for which $V > 0$ are transmitted, *vide* Fig. 2. Then b_0 is required to be zero and the $h_{0,n}$ are appropriately simplified. Sample spectra are shown in Fig. 11.

Finally we have the case in which the input wave is clipped equally on the top and bottom. A study of the condition where $b_0 = -\bar{C}$, as well as of the more general cases ($b_0 < \bar{C}$) should give some idea of the effect of saturation and cut-off on the output of an ordinary amplifier, as in practice Class A operation is often "contaminated" by some (nearly symmetrical) clipping of the larger amplitudes, due to the leveling off of the dynamic response. In all instances where the input is random noise we have departures from pure Class A performance. Here we observe that $h_{0,n}$, $n=2, 4, 6, \dots$ is zero, since then $\phi^{(n-2)}$ is an even function. In other words, of the components that contribute to the spectrum none of the even ones are generated. Figure 10 indicates some typical spectra.

Among the results should be included the correlation function for linear half-wave rectification when there is no saturation. Equation (3-4) then reduces to

$$R(t)_{b_0=0} = \frac{\beta^2 \psi}{2\pi} \left\{ 1 + \pi r/2 + \sum_{n=2,4,\dots}^{\infty} \left[\frac{(n-2)!}{2^{(n-2)/2} \left(\frac{n-2}{2}\right)!} \right]^2 r^n / n! \right\}, \quad (4-12)$$

which may be expressed in closed form:

$$R(t)_{b_0=0} = \frac{\beta^2 \psi}{2\pi} \{ (1-r^2)^{1/2} + r(\pi/2 + \sin^{-1} r) \}, \quad (4-13)$$

a result equivalent to Eq. (4.7-5) of Rice's paper.

It is instructive to examine a number of limiting cases. If the cut-off and saturation voltages are equal there should be no output, for obvious physical reasons. Setting $b_0 = C$ in (4-3)–(4-6) or more directly in (4-1) shows that this is indeed the case. Then, too, we may let C approach infinity in (4-11), when $b_0 = -C$. This corresponds to biasing so positively, i.e., $b_0 \ll 0$, that the entire disturbance is passed with negligible distortion. From (4-3) we see that $W_{d1} = C^2 \beta^2$; therefore, the power associated with the continuum is $W_c = R(0) - C^2 \beta^2 = \beta^2 \psi$, as $C \rightarrow \infty$, in other words, the input wave is completely transmitted. The spectrum is also unchanged, a fact easily verified from (4-3)–(4-6), since only h_{a1} does not vanish when $C \rightarrow \infty$, with the exception of the d.c. contribution, whose amplitude is governed by h_{a0} . Of course (in theory), infinite power must be supplied by the bias source to shift the response so that (essentially) infinitely negative noise peaks will be passed. That we obtain an infinite total power as $C \rightarrow \infty$ is clear from (4-11). Still another test may be made if we allow the bias to become greatly negative, i.e., $b_0 \gg 0$, with the condition that the saturation level C always exceed the cut-off voltage, i.e. $C > b_0 (\gg 0)$. Physically we expect a vanishing output, for only the very infrequent, large noise amplitudes will be passed, and as the bias is made more and more unfavorable, the probability of the occurrence of these becomes negligible. Examination of (4-3)–(4-6) does indeed show that $h_{a,n} \rightarrow 0$ as $C, b_0 \rightarrow \infty$, so that under such conditions there is no output.

V. THE TYPE-B QUADRATIC RECTIFIER

We may determine the Fourier transform $f(ix)_B$ of the Type-B dynamic path with the help of (1-2), by expressing $f(ix)$ as a polynomial in V , since from (1-2), $I(t)$ is a polynomial function of $V(t)$ over specified ranges of V . We start with a representation of the form

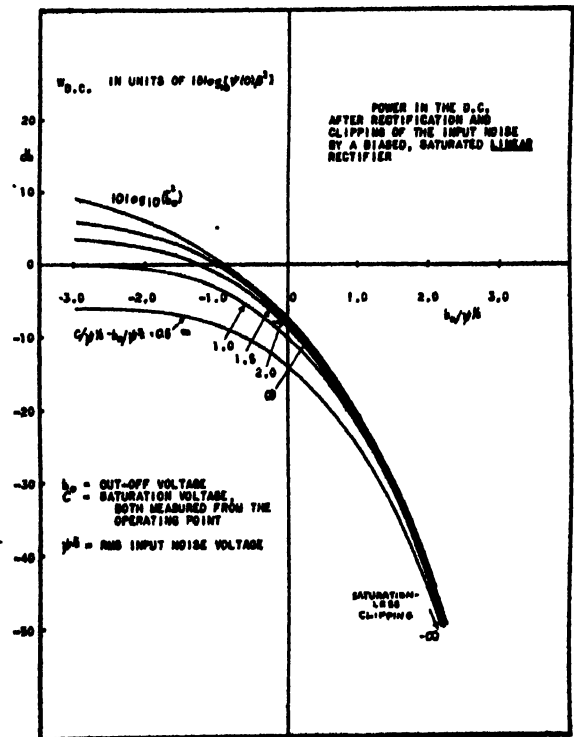
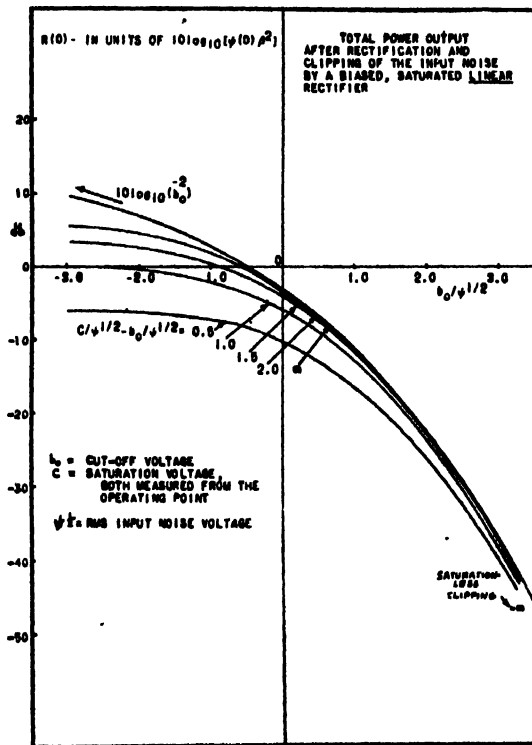


FIG. 18. Total power output of a linear rectifier for input noise of types (I–III).

FIG. 19. Power in the d.c. output of a linear rectifier for input noise of types (I–III).

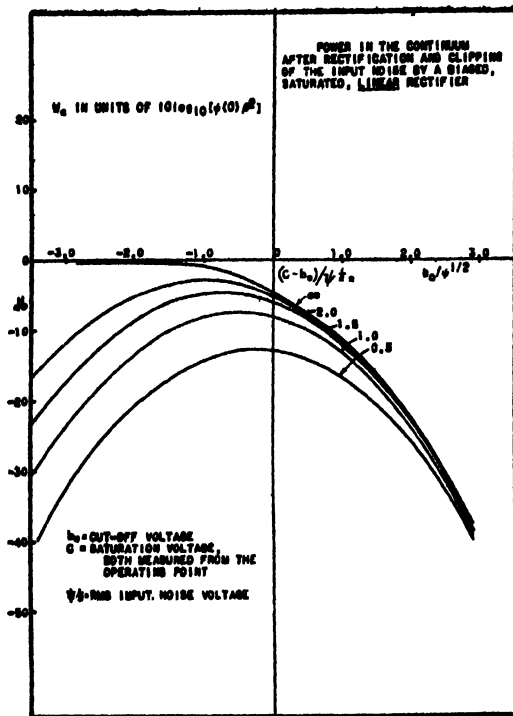


FIG. 20. Power in the continuous portion of the output of a linear rectifier, for input noise of types (I), (III).

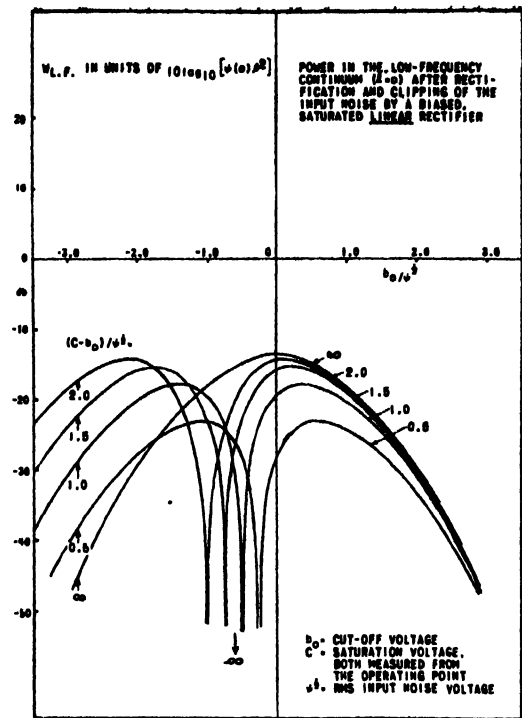


FIG. 21. Power output in the low frequency ($l=0$) continuum for type III noise entering a linear rectifier.

$$I(t) = \frac{\lambda^2}{2\pi} \int_C \left\{ \frac{2e^{i(V-b_0)z} + 2Ae^{i(V-2C+b_0)z} + 2Be^{i(V-C)z}}{(iz)^3} + \frac{De^{i(V-C)z}}{(iz)^2} + \frac{Ee^{i(V-C)z}}{(iz)} \right\} dz, \quad C \geq b_0. \quad (5-1)$$

To determine A , B , D , and E , we evaluate (5-1) for the various regions involved and match these results to the desired solution for each region. For example, when $V > 2C - b_0$, $I(t) = 2\lambda^2(C - b_0)^2$, and so

$$(V - b_0)^2 + A(V - 2C + b_0)^2 + B(V - C)^2 + D(V - C) + E = 2(C - b_0)^2, \quad (5-2)$$

from which we see that $A = 1$, $B = -2$, $D = E = 0$. The values of A , B , D , and E so obtained are easily seen to be compatible with the other regions $V < 2C - b_0$. In this manner we find for the Type-B detector that $f(iz)$ is given by

$$f(iz)_B = 2\lambda^2 \{ \exp(-ib_0z) + \exp[-i(2C - b_0)z] - 2 \exp(-iCz) \} / (iz)^3, \quad C \geq b_0. \quad (5-3)$$

Then (3-3) may be written

$$h_{0,n} = \frac{i\lambda^2}{\pi} \int_C \exp(-\psi^2/2) z^{n-3} \{ \exp(-ib_0z) + \exp[-i(2C - b_0)z] - 2 \exp(-iCz) \} dz, \quad n \geq 0. \quad (5-4)$$

Equation (5-4) may be integrated with the help of the Appendix, to give

$$h_{0,0} = -2\lambda^2 \psi \{ (\delta_0^2 + 1)(\Theta_{b_0} - 1) + 2\delta_0\phi_{b_0} + [(2C - b_0)^2 + 1][\Theta_{2C-b_0} - 1] + 2(2\bar{C} - \delta_0)\phi_{2C-b_0} - 2(\bar{C}^2 + 1)(\Theta_c - 1) + 4\phi_c \}, \quad (5-5)$$

$$h_{0,1} = -i\lambda^2 \psi^2 \{ \delta_0(\Theta_{b_0} - 1) + 2\phi_{b_0} + (2\bar{C} - \delta_0)(\Theta_{2C-b_0} - 1) + 2\phi_{2C-b_0} - 2\bar{C}(\Theta_c - 1) - 4\phi_c \}, \quad (5-6)$$

$$h_{0,2} = \lambda^2 \{ \Theta_{b_0} + \Theta_{2C-b_0} - 2\Theta_c \}, \quad (5-7)$$

$$h_{0,n} = 2i\lambda^2 (-1)^{(n-3)/2} \psi^{(2-n)/2} \{ \phi_{b_0}^{(n-3)} + \phi_{2C-b_0}^{(n-3)} - 2\phi_c^{(n-3)} \}, \quad n = 3, 5, 7, \dots, \quad (5-8)$$

$$h_{0,n} = 2\lambda^2 (-1)^{(n-2)/2} \psi^{(3-n)/2} \{ \phi_{b_0}^{(n-2)} + \phi_{2C-b_0}^{(n-2)} - 2\phi_c^{(n-2)} \}, \quad n = 4, 6, 8, \dots. \quad (5-9)$$

We have made the additional abbreviations

$$\phi_{2C-b_0}^{(n)} = \phi^{(n)}[(2C-b_0)/\psi^{\frac{1}{2}}]; \quad \Theta_{2C-b_0} = \Theta[(2C-b_0)/(2\psi)^{\frac{1}{2}}]. \quad (5-10)$$

The total output power of the Type-B quadratic device may be found from (1-2) and (3-8). We write then

$$W_r = R(0) = \frac{\lambda^4 \psi^2}{(2\pi)^{\frac{1}{2}}} \left\{ \int_{b_0}^{C/\psi^{\frac{1}{2}}} (x-b_0)^4 \exp(-x^2/2) dx + \int_{\bar{C}-b_0/\psi^{\frac{1}{2}}}^{2\bar{C}-b_0/\psi^{\frac{1}{2}}} [2(\bar{C}-b_0)^2 - (x-2\bar{C}+b_0)^2]^2 \exp(-x^2/2) dx + 4(\bar{C}-b_0)^4 \int_{x_0-b_0}^{\infty} \exp(-x^2/2) dx \right\}. \quad (5-11)$$

This we integrate to obtain finally

$$W_r = \frac{\lambda^4 \psi^2}{2} \{ P_1 \Theta_{2C-b_0} - P_2 \Theta_C - P_3 \Theta_{b_0} + 2P_4 \phi_{2C-b_0} + 2P_5 \phi_C - 2P_6 \phi_{b_0} + 4(\bar{C}-b_0)^4 \}. \quad (5-12)$$

where

$$P_1 = 16\bar{C}^3 b_0 - 28\bar{C}^2 b_0^2 + 16\bar{C} b_0^3 - 3b_0^4 + 20\bar{C}^3 - 16\bar{C}^2 + 2b_0^3 + 3, \quad (5-13a)$$

$$P_2 = 4\bar{C}^4 - 4\bar{C}^3 b_0^2 - 4b_0^3 + 20\bar{C}^3 - 16\bar{C} b_0, \quad (5-13b)$$

$$P_3 = b_0^4 + 6b_0^3 + 3, \quad (5-13c)$$

$$P_4 = 3b_0^3 - 5b_0 + 8b_0 \bar{C}^2 - 10\bar{C} b_0^2 + 10\bar{C}, \quad (5-13d)$$

$$P_5 = 4(4b_0 - 4\bar{C} + b_0^2 \bar{C} - \bar{C}^3), \quad (5-13e)$$

$$P_6 = b_0^3 + 5b_0. \quad (5-13f)$$

The power in the continuum follows when $h_0 v^2$, Eq. (5-5), is subtracted from (5-12).

It is interesting to observe for symmetrical clipping, $C=0$, $b_0 < 0$, that the Type-B rectifier yields no output in the even harmonic zones $l=0, 2, 4, \dots$, when the input noise is a narrow band. Only odd-order products contribute on the average, as is true under similar circumstances in the case of the linear detector, when $b_0 = -C$.

As before, when the cases of half-wave rectification ($b_0=0$), or symmetrical clipping (which does not obtain for Type A responses), is discussed, considerable simplification results in the expression for the output power and spectra. An important example is the half-wave, unsaturated quadratic detector. For this device we obtain the output correlation function

$$R(l)_{b_0=0} = \frac{\lambda^4 \psi^2}{2\pi} \left\{ \frac{\pi}{2} + 4r + \pi r^2 + 4 \sum_{n=3, 5, \dots}^{\infty} \left[\frac{(n-3)!}{2^{(n-3)/2} \left(\frac{n-3}{2} \right)!} \right]^2 r^n / n! \right\}, \quad (5-14)$$

which becomes, in closed form,

$$R(l)_{b_0=0} = \frac{\lambda^4 \psi^2}{2\pi} \left\{ \left(\frac{\pi}{2} + \sin^{-1} r \right) (1 + 2r^2) + 3r(1 - r^2)^{\frac{1}{2}} \right\}. \quad (5-15)$$

VI. EXTREME CLIPPING

The limiting case of extreme or "superclipping" at the "top" and "bottom" of the input is instructive as it offers an answer to such questions as whether unrestricted clipping results in an indefinitely great spectral spread, and whether the spectra so obtained appear with greatly reduced levels. Some cases of sym-

metrical clipping, including the limiting extreme of complete "slicing" of the input wave, have been considered by J. H. Van Vleck. We are now able to extend these earlier results to the general case of superclipping at any arbitrary cut-off level, i.e., for arbitrary values of b_0 . Analytically, extreme clipping means that the saturation level C approaches the cut-off b_0 so

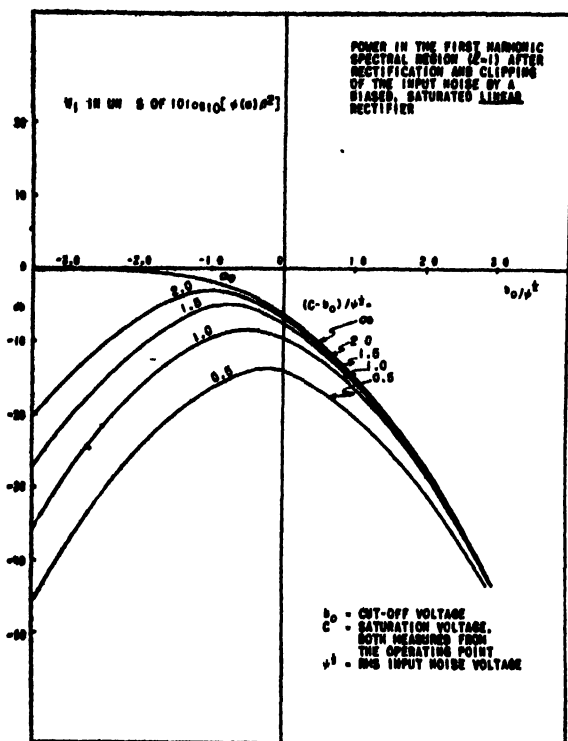


FIG. 22. Power output in the high frequency ($l=1$) continuum for type III noise entering a linear rectifier.

that the voltage difference between the two levels is infinitesimal. Then for all types of superclipping the following conditions must hold:

$$C - b_0 = \epsilon, \text{ or } C - b_0 = \xi; C > b_0, 0 < \epsilon \ll 1. \quad (6-1)$$

We may obtain the $h_{0,n}$'s from (4-3)–(4-6) by applying Taylor's theorem. Let $g(\bar{C})$ be the function whose value at $\bar{C} = b_0 + \xi$ we wish to determine. Expansion about the point b_0 gives

$$g(\bar{C}) = g(b_0) + \xi g'(b_0) + \xi^2 g''(b_0)/2 + \dots \quad (6-2)$$

It is readily verified for the linear detector that only terms in ξ are significant; the higher powers of ξ are negligible and the zeroth power ($\xi^0=1$) vanishes in $h_{0,n}$. Equations (4-3)–(4-6) are accordingly modified to

$$h_{0,0} = -\beta \xi \psi(\Theta b_0 - 1)/2, \quad (6-3)$$

$$h_{0,1} = -\beta \xi \phi b_0, \quad (6-4)$$

$$h_{0,n} = \beta (-1)^{(n+2)/2} \psi^{(1-n)/2} \phi b_0^{(n-1)}, \quad n = 2, 4, 6, \dots, \quad (6-5)$$

$$h_{0,n} = \beta \phi (-1)^{(n+3)/2} \psi^{(1-n)/2} \phi b_0^{(n-1)}, \quad n = 3, 5, 7, \dots. \quad (6-6)$$

Spectra, examples of which are shown in Figs. 7, 9, 14, 17, may be obtained from (3-1) and (3-2), with the help of (6-3)–(6-6). It can also be demonstrated for the quadratic rectifier, discussed briefly in Section V, that $h_{0,0} \sim (\Theta b_0 - 1)/2$ and $h_{0,n} \sim \phi b_0^{(n-1)}$, in line with the results above, the proportionality constant $\beta \xi$ being replaced by $2\lambda^2 \xi^2$, indicating that the shape of the output spectrum is independent of the dynamic $i_p - e_p$ path if ξ is taken sufficiently small, a reasonable conclusion, based on the observation that the output is now essentially a square wave with but two possible amplitudes, determined by the limit of the transmission width and not by the dynamic path between those limits.

The correlation function of the output in cases of extreme clipping is for any input distribution,

$$R(t) = \xi^2 \psi \beta^2 \left\{ \left(\frac{1 - \Theta b_0}{2} \right)^2 + \sum_{n=1}^{\infty} [\phi b_0^{(n-1)}]^2 r(t)^n / n! \right\}. \quad (6-7)$$

The total power may be obtained as before by setting $t=0$ in (6-7), or by applying (4-10). We have

$$W_r = R(0) = \xi^2 \beta^2 (1 - \Theta b_0)/2. \quad (6-8)$$

The power in the continuum is therefore

$$W_c = R(0) - h_{0,0}^2 = \xi^2 \beta^2 (1 - \Theta b_0^2)/4. \quad (6-9)$$

If we specialize our results to symmetrical clipping we need only multiply (6-7)–(6-9) by 4. This follows because the effective amplitude of the disturbance is doubled, and hence the power is increased fourfold. When the incoming wave undergoes half-wave clipping and rectification, b_0 is zero, and Eq. (6-7) reduces to

$$R(t)_{b_0=0} = \xi^2 \beta^2 \left\{ \frac{1}{4} + \frac{1}{2\pi} \sum_{n=1,3,\dots}^{\infty} \left[(n-1)! / 2^{(n-1)/2} \left(\frac{n-1}{2} \right)! \right]^2 r(t)^n / n! \right\}, \quad (6-10)$$

which may be written as

$$R(t)_{b_0=0} = \xi^2 \beta^2 \left\{ \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} r \right\} = \xi^2 \beta^2 \left\{ \frac{1}{4} + \frac{r}{2\pi} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; r^2\right) \right\}, \quad (6-11)$$

a result similar to that obtained by Van Vleck. A rather interesting consequence of extreme clipping is that zero cut-off voltage ($b_0 = 0$), gives the greatest actual power in the continuum ($\neq$ non-vanishing), while for $b_0 \rightarrow \infty$, i.e., very negative bias, a greater *percentage* of the power goes into the continuum.

VII. RESULTS AND CONCLUSIONS

In the preceding sections we have outlined the analysis from which the mean power and spectral output of biased and saturated linear or quadratic rectifiers may be determined when the incoming wave is random noise. Since much of our work appears specifically in the figures, it is with them principally in mind that we discuss the results. The transmission width $C - b_0$, or $2(C - b_0)$, is treated as constant, which it is for resistive loads, so that as the amount of incoming noise is varied, the *relative* transmission width $(C - b_0)/\psi^{\frac{1}{2}}$, or $2(C - b_0)/\psi^{\frac{1}{2}}$, takes on different values. This corresponds physically to the operation of a given tube with various applied grid voltages and bias. The effects of clipping on the output spectra have been compared by normalizing the maximum spectral level of the output to unity. The normalization is accomplished after multiplication by the appropriate quantity, here denoted by A . For example, in Fig. 8 when $(C - b_0)/\psi^{\frac{1}{2}} = 0.5$ and $b_0/\psi^{\frac{1}{2}} = 1.5$, ($\sigma = 2.0$), A is seen to be 612 or its equivalent 27.9 db, where the db are determined from $10 \log_{10} A$. To obtain the absolute spectral level we have only to divide the normalized value by A .

Extensive calculations and figures corresponding to Figs. 5-23 have also been completed for the Type A and Type B quadratic rectifiers. It has been found that the results in the latter cases are qualitatively similar to those for the linear rectifier, as may be seen in a few special instances, cf. Figs. 24 and 25, on comparison with Figs. 5 and 18. The general shape of the spectra is much the same, and the various output power distributions, analogous to Figs. 18-22 are similar. A comparison of absolute levels cannot in general be made, however, as it requires a specific knowledge of the tube constants β and γ , λ , which in turn are functions of the operating conditions. Because of general similarity of the

results for the three types of detector, we have omitted the detailed analyses and the figures in the square-law cases, but these are available on request.

(a) Spectra

The first observation to be made from Figs. 5-17 is that clipping, whether at the "top" or "bottom" of the input wave, *always* results in a spectral spread.* This may be appreciated in a qualitative fashion from the following simple example. Consider first the spectrum of a sine wave. It will contain just one component, the original wave. Now clip the wave, at the top or bottom, or at both places. The signal is distorted and has "corners" where the clipping took place. When the disturbance is analyzed, we find a large number of new, high frequency terms; these are necessary to give the "squared-off" shape of the wave. Physically, a fraction of the input power is lost, since part of the signal is removed, and the rest of the input power is distributed among the higher harmonics generated by clipping, the precise redistribution depending on the wave's amplitude, the tube characteristic, etc. The above example is easily extended to noise with its infinite number of components.

The curves of Fig. 8 also show that the spectral spread is greatest when clipping occurs because of the tube cut-off alone and not on account of saturation as well. There are a few exceptions to this, which occur only for narrow-band input noise, as Figs. 13 and 14 show. Note that increasing the relative transmission width $(C - b_0)/\psi^{\frac{1}{2}}$ yields clipping which arises principally from cut-off, for a given value of b_0 , since overloading effects then tend to disappear. When the rectified noise output is passed into amplifiers the level alone is altered and the character of the noise is preserved, *provided* the amplification is a purely linear process. In this connection we observe that filters, of which amplifiers are an example, discriminate on a frequency basis, and affect the instantaneous amplitude of the wave only indirectly as they respond or fail to respond to the various components. Filtering tends to

* That clipping always spreads the spectrum in the case of noise may be proved by summing the limiting forms of $k_{0,n}^2 c_{0,n}(f)$ as $b_0 \rightarrow \infty$, and comparing the normalized sum with that for finite values of b_0 .

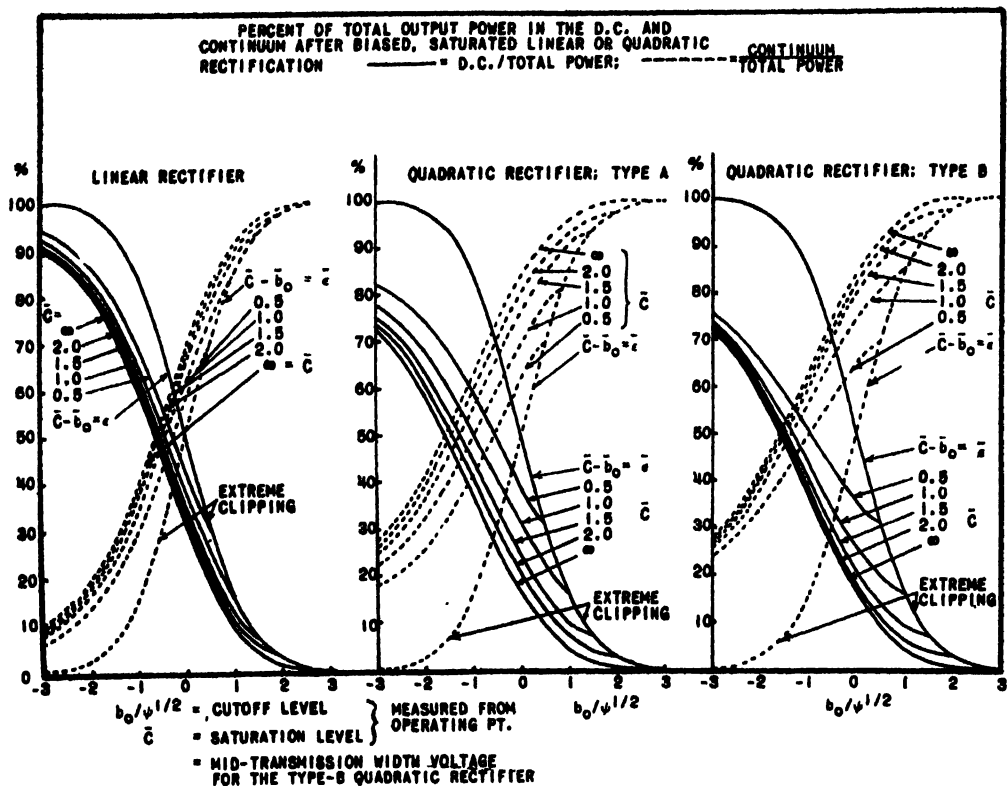


FIG. 23. Fraction of the total power in the d.c. and continuous portions of the output for the linear, Type-A, and Type-B quadratic detectors.

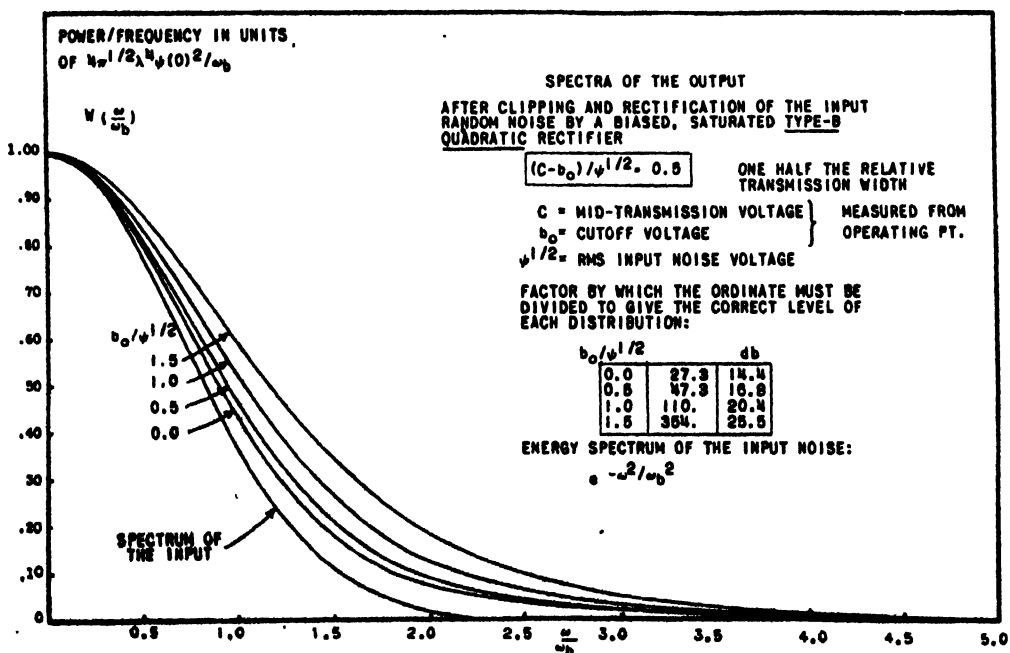


FIG. 24. Output spectra for broad band input noise (I) from a biased and saturated Type B quadratic detector.

reduce the spread in the spectrum of clipped noise and when the band width of the filter is narrow compared with the spectrum, randomness is restored and the effects of clipping are no longer apparent.

From Fig. 12 for the low frequency output of narrow band noise ($l=0$) it appears that when clipping occurs near the center of the wave, i.e. $b_0=0$, the spread is greatest, although the level is well down. In the case of the first harmonic output zone, Fig. 15, $l=1$, corresponding to a band centered about f_0 , the reverse seems to be true, and the greatest spread takes place for the larger values of cut-off b_0 , although this spread is not so pronounced. The reason for this may be appreciated by reference to Eq. (3-23) which shows that when $l=1$, the leading term is generated by taking $n=1$; on the other hand, for the low frequency continuum $l=0$, and the leading term starts with $n=2$; in the former instance we have the spectral contribution $\exp(-\omega^2/\omega_b^2)$, in the latter, $\exp(-\omega^2/2\omega_b^2)$, which suggests why the latter in many cases yields the greater spread. We may remark also that in the neighborhood of spectral "tails," well away from the center of the distribution, the spread is always greatest for extreme clipping.*

Figure 11 illustrates the spectral spread after half-wave rectification with saturation, and shows the effects either of decreasing the input-noise voltage or of using tubes whose dynamic paths cover a progressively wider range of voltages. In either case, as the relative transmission width is increased, greater spectral spread obtains and at a higher level, until the limit of saturationless clipping is reached. The level rises because more of the incoming wave is transmitted, and again it is seen that clipping alone due to the tube cut-off results in the greatest spectral spread and at the highest (relative) output level.*

* At first glance this may appear odd, since we would expect the additional clipping from overload to spread the spectrum even more. A comparison of the correlation functions for extreme clipping when $b_0=0$, Eq. (6-10), and for half-wave rectification without saturation ($b_0=0$, $C=\infty$), Eq. (4-12), shows that $R(t)_{ext} \sim r + r^2/6 + 3r^4/40 + \dots$, while $R(t)_{no sat} \sim r + r^2/\pi + r^4/12\pi + \dots$. It is the greater magnitude of the second term, r^2/π vs. $r^2/6$, that accounts for the greater spread in the instance of no saturation. It should be observed, however, that a comparison of spectra in the neighborhood of the "tail," well away from the center

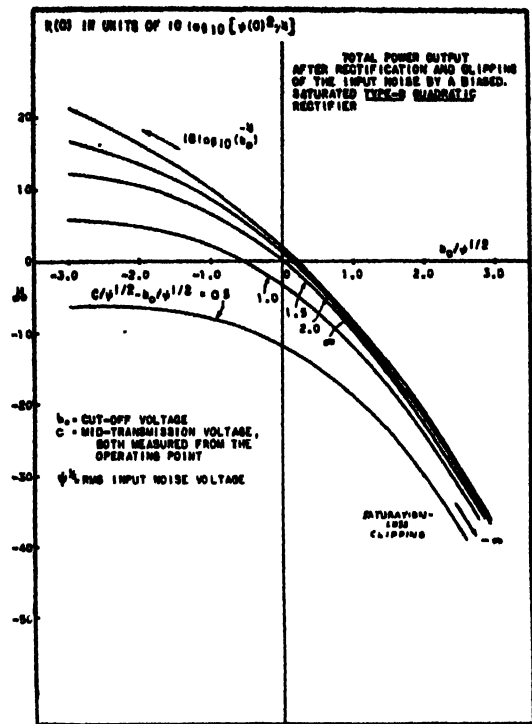


FIG. 25. Total power output of a Type B quadratic rectifier for input noise of classes I-III.

Examples of symmetrical clipping are shown in Fig. 10, in which the rectifier is biased so that the operating point lies half-way between cut-off and saturation, i.e., $b_0=C$. For broad band and semi-broad band noise little spreading of the spectrum results, even when the relative transmission width is greatly decreased, but when the incoming disturbance is narrow band (III), the low frequency ($l=0$) output vanishes. In fact, all even harmonics $l=0, 2, 4, \dots$ are missing when the clipping is symmetrical. This effect may also be observed in Fig. 21 where the output power goes to zero. As we have mentioned before, amplification in practice nearly always results in approximately symmetrical clipping at both top and bottom of the wave; these results show that the effect is small for input spectra of types I and II, but may be serious in certain output zones of class III waves.

Figures 7, 9, 14, and 17 illustrate instances for which the relative transmission width is in-

of the band, indicates that the reverse is true. A similar comparison may be made for the quadratic detector with the help of (5-14).

infinitesimally small, producing extreme or "superclipping." We see that superclipping spreads the spectrum most when the dynamic path is centered well away from the operating point, i.e., $|b_0| \gg 0$, where the curvature is greatest. (An exception to this is the case $l=0$, Fig. 14.) The spectral level, of course, is extremely low.

(b) Power

As we have demonstrated in Section 3, the power output of our non-linear device, unlike the spectrum of the output, does not depend on the spectrum of the input disturbance, but only on the root-mean-square input-noise voltage. Figure 19 shows the d.c. output power as a function of cut-off voltage, measured from the operating point. As b_0 becomes more positive less of the wave is transmitted and hence the direct current drops and approaches the limiting value for saturationless clipping, which depends only on b_0 and not on the transmission width $(C-b_0)$. On the other hand, when b_0 is negative, the positive part of the incoming disturbance is progressively more and more limited because of saturation. Then the d.c. output rises and approaches a fixed value asymptotically, since it cannot exceed some finite value which depends on the transmission width of the rectifier. For no overloading ($C \rightarrow \infty$), this finite limit is removed and represents chiefly the voltage supplied by the tube bias, which is needed to shift the operating point further into the transmission region of the dynamic path.

The powers in the continuum appear in Figs. 20, 21, and 22. Here it is interesting to observe that for symmetrical clipping in the cases of class I and II input spectra, the continuum level is highest, falling off for values of $|b_0|$ in excess of $b_0 = -C$. This may be explained by the fact that when the transmission width is centered about the incoming wave more of the wave is effectively passed than if only the tops or bottoms are transmitted. When disturbances having class III spectra are considered, this is seen to be true for the first and higher odd harmonic regions $l=1, 3, 5, \dots$, vide Fig. 22, but for values of $b_0 = -C$, there is no low frequency or for that matter, no even-harmonic output, because, with the exception of a d.c. component, only odd-order modulation products are gener-

ated. Curves of total output power are given in Fig. 18. Figure 23 shows the percentage distribution of the total output power that appears in the continuum and in the d.c. component. The quadratic responses, it may be noted, yield a somewhat greater fraction of continuum power than does the linear detector, and of these the Type B characteristic yields the most, for a given cut-off and transmission width.

We expressly wish to thank Professor J. H. Van Vleck for his encouragement in the preparation of this manuscript, Mr. S. O. Rice, with whom we have had many friendly and stimulating discussions on these topics, and Miss Eleanor Pressly, who performed most of the computations.

APPENDIX

In Section III we gave the results of the evaluation of (4-2) for the biased and saturated linear detector. A similar procedure is used to evaluate (5-4) for the quadratic rectifier.

Let us consider only the first term in the integrand of (4-2), since the second may be treated in a similar fashion. When $n=0$, the imaginary part contributes only through the pole at the origin, i.e.,

$$(h_0)_0 = \frac{-\beta}{2\pi} \left\{ \int_C \cos b_0 z \exp(-\psi z^2/2) dz/z^2 + \pi b_0 \right\}, \quad (\text{A-1})$$

since the effective contour is a semicircle. The real part, however, is somewhat more difficult to evaluate. By adding $-1/z^2$ to the integrand we may remove the singularity and hence replace C by a path entirely along the real axis. It is easily shown that $\int_C z^{-2} dz$ vanishes. We now have an integral of the form

$$I_1 = \int_{-\infty}^{\infty} [\cos b_0 z \exp(-\psi z^2/2) - 1] dz/z^2. \quad (\text{A-2})$$

Differentiating (A-2) with respect to b_0 we obtain

$$\frac{\partial I_1}{\partial b_0} = -2 \int_0^{\infty} \sin b_0 z \exp(-\psi z^2/2) dz/z \\ = -\pi \Theta(b_0/(2\psi)^{1/2}), \quad (\text{A-3})$$

where Θ is defined in (4-8). Integration of (A-3) by parts yields

$$I_1(b_0) - I_1(0) = -\pi \{ b_0 \Theta(b_0/(2\psi)^{1/2}) (\exp(-b_0^2/2\psi) - 1) \}. \quad (\text{A-4})$$

Now from (A-2) we observe that

$$I_1(0) = \int_{-\infty}^{\infty} [\exp(-\psi z^2/2) - 1] dz/z^2; \quad (\text{A-5})$$

hence we may write

$$\frac{\partial I_1(0)}{\partial \psi} = - \int_{-\infty}^{\infty} \exp(-\psi z^2/2) dz = -(\pi/2\psi)^{1/2}. \quad (\text{A-6})$$

Finally we have

$$I_1(0) = -(\pi/2\psi)^{1/2}. \quad (\text{A-7})$$

The constant of integration is zero here, since for $\psi=0$ $I_1(0)=0$. From (A-4) and (A-7) it follows that

$$(h_0)_0 = \beta \{ b_0 (\Theta(b_0/(2\psi)^{1/2}) - 1) / 2 + \psi^{1/2} \psi_0 \}, \quad (\text{A-8})$$

with a similar result when b_0 is replaced by C . Equation (4-3) may be obtained from (4-2) and the above. For $n=1$, we have

$$(h_{1,1})_{b_0} = -\frac{\beta}{2\pi} \left[\pi i - i \int_{-\infty}^{\infty} \sin b_0 s \exp(-\psi s^2/2) ds/s \right] \\ = \beta i (\Theta b_0 - 1)/2. \quad (\text{A-9})$$

from which (4-4) is established in a similar fashion. When $h \geq 2$, the contour C is replaced by a path along the real axis. Then to determine $h_{n,n}$ we need the following results, which are readily established by differentiation with respect to λ :

$$\int_0^{\infty} s^{2p} \cos \lambda s \exp(-\psi s^2/2) ds \\ = (\pi/2)^{1/2} (-1)^p \psi^{-p-1/2} H_{2p}(\lambda/\psi^{1/2}) \exp(-\lambda^2/2\psi) \quad (\text{A-10}) \\ = \pi (-1)^p \psi^{-p-1/2} \phi^{(2p)}(\lambda/\psi^{1/2}),$$

and

$$\int_0^{\infty} s^{2p-1} \sin \lambda s \exp(-\psi s^2/2) ds \\ = (\pi/2)^{1/2} (-1)^p \psi^{-p-1/2} H_{2p-1}(\lambda/\psi^{1/2}) \exp(-\lambda^2/2\psi), \quad (\text{A-11}) \\ = \pi (-1)^p \psi^{-p-1/2} \phi^{(2p-1)}(\lambda/\psi^{1/2}), \quad p=0, 1, 2, \dots$$

where $H_j(x_0)$ is an Hermitian polynomial of the j th order,¹³ cf. Eqs. (4-9). Here we note also that¹⁴

$$H_{-1}(x_0) = (2\pi)^{1/2} \phi^{(-1)}(x_0) \\ = (\pi/2)^{1/2} \exp(+x_0^2/2) \Theta(x_0/2^{1/2}), \quad (\text{A-12})$$

and that

$$H_{2p}(0) = (-1)^p (2p)! / 2^p p! \\ H_{2p-1}(0) = 0, \quad p=0, 1, 2, \dots \quad (\text{A-13})$$

The explicit form for $\phi^{(j)}(x_0)$ is useful, namely

$$\phi^{(j)}(x_0) = \exp(-x_0^2/2) \sum_{p=0}^{j/2} \frac{(j-1)!}{p!} (-1)^{p+1} j! x_0^{j-2p} / \\ p! (j-2p)! 2^p (\pi)^{1/2}. \quad (\text{A-14})$$

The functions $\phi^{(j)}$ satisfy the recurrence relation

$$\phi^{(j+1)}(x_0) = -[x_0 \phi^{(j)}(x_0) + j \phi^{(j-1)}(x_0)], \\ j=1, 2, \dots \quad (\text{A-15})$$

¹³ See, for example, T. C. Fry, *Probability and Its Engineering Uses* (D. Van Nostrand Co., Inc., 1928), pp. 251-254, and for tables of $\phi^{(0)} - \phi^{(6)}$ consult pages 456, 457.

¹⁴ B. O. Pierce, *A Short Table of Integrals* (Ginn and Company, New York, 1929), contains a table of Θ .

Calendar of Meetings

October

25 American Mathematical Society, New York, New York

November

1-2 Society of Rheology, New York, New York

2 American Physical Society (New York State Section), Saratoga Springs, New York

7-9 Conference on Electrical Insulation, Division of Engineering and Industrial Research, National Research Council, Baltimore, Maryland

14-16 Acoustical Society of America, Chicago, Illinois

29-30 American Physical Society, Minneapolis, Minnesota

December

2-6 American Society of Mechanical Engineers, New York, New York

5-7 Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction, Pittsburgh, Pennsylvania (Joint Meeting)

17 Institute of Aeronautical Sciences, New York, New York

26-28 Geological Society of America, Chicago, Illinois

January

23-25 American Physical Society, New York, New York

The Flow of Gases in Pipes at Low Pressures

GORDON P. BROWN AND ALBERT DiNARDO
National Research Corporation, Boston, Massachusetts

AND
GEORGE K. CHENG AND THOMAS K. SHERWOOD
Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 9, 1946)

Data are presented on the flow of air at low pressures in copper pipes of radius 0.795 and 1.30 cm; iron pipes of radius 2.64, 5.12, and 10.1 cm; and for hydrogen in a copper pipe of radius 1.30 cm. Data are also given for the resistance to flow of air in 1½- and 3-inch nominal pipe size elbows, for 1½-inch Kinney bellows valves, and for short sections of 1-inch nominal pipe size iron pipe connecting large gas chambers. The data on straight iron and copper pipes, together with the data from the literature on glass capillaries, are correlated by the introduction of a correction factor F in Poiseuille's equation. F is correlated graphically in terms of a dimensionless group representing the ratio of mean free path to pipe radius, with separate curves for iron pipe and for copper and glass pipes or capillaries.

INTRODUCTION

INDUSTRIAL processes operating at high pressure have been common for a generation, but it is only during the last several years that the other extreme of operating pressures has come to be used in a variety of large scale operations. Pressures as low as a few microns are employed in the manufacture of electronic equipment, production of magnesium metal from dolomite, the drying of blood plasma and of penicillin, the concentration of vitamins, and the coating of lenses used in military optical equipment.¹ Moderately low pressures are being used in concentration of fruit juices and other food products.

The necessary vacuum is obtained by the use of mechanical pumps, multi-stage steam jets, or large oil diffusion pumps. No matter how good the vacuum at the pump, it is evident that the pressure in the processing equipment may be unnecessarily high and perhaps excessive if the piping and connections are not designed to handle the gas or vapor flow with an economically small pressure loss due to friction. The proper design of conduits for gas flow at low pressures introduces new problems. At ordinary pressures it is customary to employ the combination of the Fanning equation and a friction factor graph if the flow is turbulent (Reynolds number greater than about 2100) and to use Poiseuille's law if the flow is viscous or laminar (Reynolds number

below about 1000). At low pressures (below 1.0 mm Hg) the Reynolds number is probably always so low as to indicate laminar flow, but Poiseuille's law does not apply since the effective viscosity, even in industrial pressure ranges, may be less than one percent of the gas viscosities as tabulated in the standard handbooks.

FLOW EQUATIONS—STREAMLINE FLOW

The differential and integrated forms of Poiseuille's law for round tubes may be written

$$Q = -\frac{\pi g_0 R^4 p}{8\mu} \frac{dp}{dL} = \frac{\pi g_0 R^4}{16\mu L} (p_1^2 - p_2^2). \quad (1)$$

The derivation of this equation, frequently quoted in textbooks, is based on the assumption that the coefficient of viscosity, μ , is a constant and that the velocity of gas in contact with the tube wall is zero. Gas viscosities, as predicted by the kinetic theory, are essentially constant over a wide range of moderate and high pressures. Consequently, the failure of the equation at low pressures is explained as caused by the failure of the assumption of no slip at the wall rather than on the basis of a variable viscosity.

SLIP FLOW

The derivation of Poiseuille's equation has been repeated by Kundt and Warburg^{2,3} with

² A. Kundt and E. Warburg, *Pogg. Ann.* 155, 337, 525 (1875).

³ E. Warburg, *Wied. Ann.* 159, 399 (1876).

¹ Anonymous, *Fortune* 20, No. 2, 147 (August, 1944).

TABLE I. Systems of units.

	Consistent c.g.s. technical	Consistent English	Micron c.g.s.	Micron English
Q	$\frac{\text{cm}^3}{\text{sec.}} \times \frac{\text{g.}}{\text{cm}^3}$	$\frac{\text{ft.}^3}{\text{sec.}} \times \frac{\text{lb.}}{\text{ft.}^3}$	$\frac{\text{micron}}{\text{sec.}} \times \text{cm}^3$	$\frac{\text{micron} \times \text{ft.}^3}{\text{sec.}}$
μ^*	$\frac{\text{g.}}{\text{sec.} \times \text{cm}} (\text{poises})$	$\frac{\text{lb.}}{\text{sec.} \times \text{ft.}}$	poises	$\frac{\text{lb.}}{\text{sec.} \times \text{ft.}}$
Rg	84800	1543	6.24×10^{-7}	5.56×10^6
R, L, λ	cm	ft.	cm	ft.
p_1, p_2, p_m	g/cm ²	lb./ft. ²	microns	microns
ρ_1	$\frac{\text{g.}}{\text{cm}^3} \times \frac{\text{cm}^3}{\text{g.}}$	$\frac{\text{lb.}}{\text{ft.}^3} \times \frac{\text{ft.}^3}{\text{lb.}}$	$\frac{\text{g.}}{\text{cm}^3} \times \text{microns}$	$\frac{\text{lb.}}{\text{ft.}^3} \times \text{microns}$
f, π	No units			
g_0	981	32.2	1.33	0.0896

* Viscosity in lb./sec.(ft.) = 0.0672 $\times$ viscosity in poises = 0.000672 $\times$ viscosity in centipoises.

the modification that the velocity of the gas in contact with the wall is taken as u_0 instead of zero when the mean free path is comparable with the tube radius. The resulting tangential force F' on the wall is assumed to be proportional to u_0 and to the surface area

$$F' = k u_0 A_w, \quad (2)$$

where the proportionality constant k is termed the coefficient of external friction. Balancing the forces on a differential length dL of tube,

$$-\pi R^2 g_c dp = F' = 2\pi R k u_0 dL, \quad (3)$$

whence

$$u_0 = \frac{R g_c}{2k} \frac{dp}{dL}. \quad (4)$$

This last is used as one limit of u in the usual integration to obtain:

$$Q = \frac{\pi g_c R^4}{16\mu L} (p_1^2 - p_2^2) \left(1 + \frac{4\mu}{kR} \right). \quad (5)$$

Maxwell⁴ suggested that a fraction f of the gas molecules striking the walls of a vessel would be diffusely reflected, and that the fraction $1-f$ would be specularly reflected. With this concept, it can be shown by the kinetic theory that,

$$\frac{\mu}{k} = \left(\frac{2}{f} - 1 \right) \lambda_m, \quad (6)$$

⁴ J. C. Maxwell, *The Scientific Papers of James Clerk Maxwell* (University Press, Cambridge, 1890), Volume 2, p. 708.

which may be introduced into (5) to give,

$$Q = \frac{\pi g_c R^4}{16\mu L} (p_1^2 - p_2^2) \left[1 + 4 \left(\frac{2}{f} - 1 \right) \frac{\lambda_m}{R} \right], \quad (7)$$

where λ_m is the mean free path of the gas molecules at the mean pressure. Since the coefficient of external friction, k , has been found⁵ to be proportional to gas pressure, its value and that of λ should be evaluated at the mean gas pressure. From the kinetic theory and Maxwell's distribution of molecular velocities, the mean free path can be shown to be given by

$$\lambda_m = \left(\frac{\pi}{2\rho_1 g_c} \right)^{\frac{1}{2}} \cdot \frac{\mu}{p_m},$$

where ρ_1 is the gas density at unit pressure, $M/R_g T$. Thus, (7) becomes

$$Q = \frac{\pi g_c R^4}{16\mu L} (p_1^2 - p_2^2) \left[1 + 4 \left(\frac{\pi}{2} \right)^{\frac{1}{2}} \left(\frac{2}{f} - 1 \right) X \right], \quad (8)$$

where X represents the group $\mu/p_m R (\rho_1 g_c)^{\frac{1}{2}}$. Inspection of (7) shows that this equation for "slip flow" is essentially Poiseuille's equation (1) multiplied by a correction factor which becomes more important as the ratio of mean free path to radius increases and as f becomes small (smooth walls). Equation (8) is a restatement of (7) in terms of more directly useful variables.

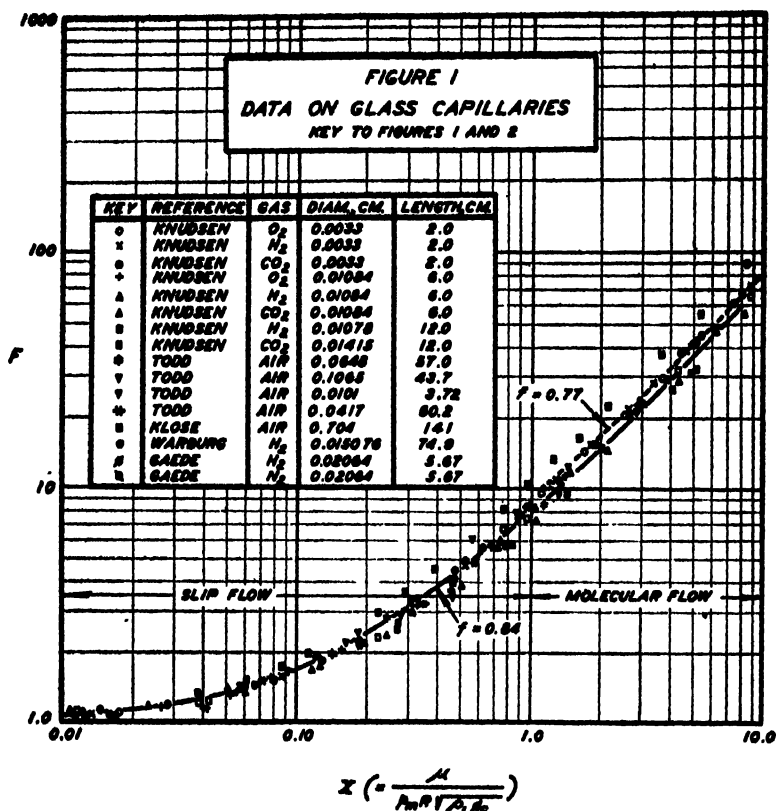
MOLECULAR FLOW

At very low pressures the internal friction in the moving gas becomes small because the mean free path of the gas molecules becomes large in relation to the dimensions of the conduit. Under these conditions the collisions between molecules are rare as compared with the frequency of collisions with the walls. By employing the kinetic theory in a balance of momentum transfer and pressure difference in a pipe section, there is obtained⁶

$$Q = \left(\frac{\pi}{2} \right)^{\frac{1}{2}} \left(\frac{2}{f} - 1 \right) \frac{R^3 g_c}{L} \frac{(p_1 - p_2)}{(\rho_1 g_c)^{\frac{1}{2}}}. \quad (9)^6$$

⁶ R. A. Millikan, *Phys. Rev.* 21, 217 (1923).

⁶ M. Knudsen [*Ann. d. Physik* 28, 75, 999 (1909)] taking f as unity, obtained the same form with a constant 70 percent greater.



The nature of the motion of the gas at very low pressures is termed "molecular flow" and Eq. (9) is a quantitative expression for the rate of molecular flow.

Inspection of the "slip flow" Eq. (8) reveals that for very low pressures the second of the two additive terms becomes large in relation to unity, and the equation reduces to (9). Equation (8) is more than an equation for the slip flow region; at high pressures it reduces to Poiseuille's equation (1), and at very low pressures it reduces to the molecular flow Eq. (9).

ENTRANCE AND EXIT LOSSES

For flow of gases through an orifice at very low pressures ("molecular effusion") Knudsen⁶ has derived the following equation from the kinetic theory

$$Q = \left(\frac{\pi}{2 \rho_1 g_0} \right)^{\frac{1}{2}} R^2 (p_1 - p_2) g_0. \quad (10)$$

Comparison of this with Eq. (9) for molecular flow indicates that molecular flow in both orifices

and tubes can be expressed by the relation,

$$Q = K \left(\frac{\pi}{2 \rho_1 g_0} \right)^{\frac{1}{2}} R^2 (p_1 - p_2) g_0, \quad (11)$$

where $K=1$ for $L/R=0$ and approaches

$$K = \frac{\pi}{2} \left(\frac{2}{f} - 1 \right) \frac{R}{L}$$

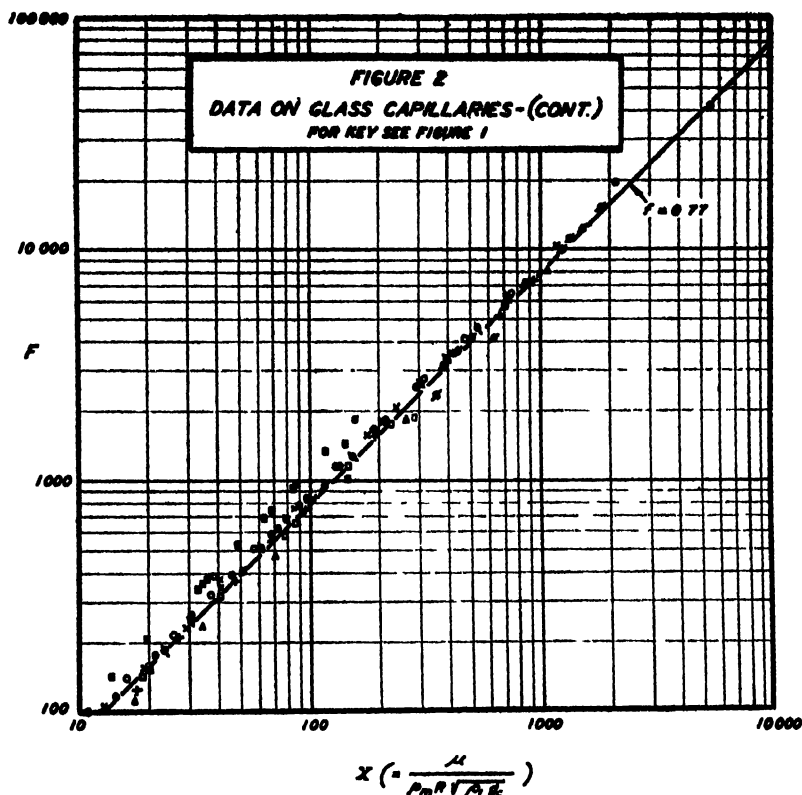
as L/R approaches infinity. Clausing⁷ gives values of K as a function of R/L , varying from 1.0 to

$$\frac{8}{3} \frac{R}{L}.$$

The latter limit corresponds to a constant value of f of 0.74.

In streamline flow the entrance loss may be taken as equal to the velocity head, u^2/g_0 . A graph of pressure drop vs. Q for an orifice or for the entrance to a tube may be represented in part by two straight lines on logarithmic co-

⁷ P. Clausing, Ann. d. Physik [5F] 12, 961 (1932).



ordinates. At low pressures the relation is indicated by (11) and the slope is unity. In streamline flow

$$Q = \pi R^2 p_2 \left(\frac{R_0 T (p_1 - p_2) g_0}{p_m M} \right)^{\frac{1}{2}} \quad (12)$$

and there is a family of lines having slopes of 2. The data cross from the first line to one of the second type as the pressure increases, and there is some indication⁸ that the experimental points pass through an inflection and a minimum in so doing.

UNITS

The various equations given in the preceding sections are valid in any consistent set of technical units; if g_0 is taken as unity, they are valid in absolute units. The common mixture of pressure in microns Hg with English units is permissible if an appropriate value of g_0 is employed.

⁸ G. K. Cheng, Sc.D. Thesis in Chemical Engineering, Massachusetts Institute of Technology, 1945.

The proper use of the four technical systems is indicated in Table I.

SUMMARY OF THEORY

Although the flow of gases at low pressures may be described as streamline flow, slip flow, or molecular flow, the single Eq. (8) is seen to represent the theory for all three regions. Alternatively, Poiseuille's equation (1) may be employed, with the right-hand side multiplied by a correction factor F :

$$S = \frac{Q}{(p_1 - p_2)} = \frac{F \pi R^4 p_m g_0}{8 \mu L}, \quad (13)$$

where

$$F = 1 + 4 \left(\frac{\pi}{2} \right)^{\frac{1}{2}} (2/f - 1) X. \quad (14)$$

Here μ is the viscosity of the gas at atmospheric pressure, and the factor F may be looked upon as the ratio of μ to the effective viscosity applicable to Poiseuille's equation at low pressures.

TABLE II. Experimental data on flow of gases in copper and iron pipes. (p_1 and p_2 represent upstream and downstream pressures in microns; U is the speed in micron cu. ft. per sec. per micron.)

Run	p_1	p_2	U	F	$\frac{\mu}{p_1 R (\rho_1 \rho_2)^{1/2}}$	Run	p_1	p_2	U	F	$\frac{\mu}{p_1 R (\rho_1 \rho_2)^{1/2}}$
A. Copper pipe, $R=1.30$ cm, $L=332$ cm, air, using 4-in. oil diffusion pump and Megavac.						C. Copper pipe, $R=0.795$ cm, $L=127$ cm; air, using N.R.C. oil diffusion pump H-6 and 50 cfm Kinney mechanical pump.					
1	4	1	0.020	9.48	1.26	1	470	130	0.099	1.07	0.0172
2	5	1	0.0186	7.28	1.05	2	370	100	0.083	1.14	0.022
3	6	2	0.0197	5.78	0.79	3	270	70	0.061	1.15	0.030
4	7	2	0.0195	5.09	0.70	4	198	48	0.045	1.18	0.042
5	8	3	0.025	5.25	0.56	5	148	34	0.037	1.31	0.056
6	10	4	0.022	3.75	0.45	6	84	15	0.023	1.48	0.103
7	11	5	0.026	3.74	0.39	7	48	7	0.0168	1.93	0.184
8	12	4	0.026	3.78	0.39	8	29	4	0.0140	2.73	0.31
9	14	5	0.027	3.31	0.33	9	14	2	0.0126	5.06	0.64
10	18	8	0.029	2.62	0.24	10	7	1	0.0120	9.7	1.29
11	20	16	0.028	1.85	0.185	D. Iron pipe, $R=2.64$ cm, $L=274$ cm; air, using N.R.C. oil diffusion pump H-6, and 50 cfm Kinney mechanical pump.					
12	28	23	0.038	1.74	0.124	1	1.35	0.25	0.136	9.45	1.94
13	30	25	0.043	1.85	0.115	2	4.00	0.80	0.159	3.78	0.65
14	38	32	0.049	1.66	0.090	3	7.80	1.60	0.180	2.18	0.33
15	53	46	0.066	1.56	0.064	4	23.20	6.50	0.34	1.31	0.105
16	54	47	0.066	1.45	0.063	5	14.60	3.80	0.26	1.61	0.168
17	64	55	0.073	1.44	0.053	6	41.20	13.0	0.54	1.14	0.057
18	70	61	0.075	1.35	0.048	7	59.9	22.4	0.85	1.17	0.038
19	77	68	0.082	1.33	0.044	8	102.0	57.0	1.36	0.93	0.0195
20	79	69	0.088	1.39	0.043	9	110.0	66.5	1.37	0.89	0.0175
21	90	78	0.087	1.21	0.038	10	136.0	100.0	2.12	1.02	0.0132
22	90	79	0.090	1.25	0.037	11	187	150	2.72	0.92	0.0092
23	120	106	0.114	1.19	0.028	12	248	211	3.79	0.94	0.0068
24	143	128	0.132	1.14	0.023	13	320	290	4.81	0.90	0.0051
25	159	142	0.143	1.12	0.021	14	385	352	5.94	0.89	0.0041
26	162	146	0.151	1.15	0.021	15	447	412	6.69	0.88	0.0036
27	167	150	0.152	1.13	0.020	16	558	528	9.91	1.04	0.0029
28	188	170	0.166	1.09	0.0186	17	620	588	10.6	1.00	0.0026
29	223	203	0.190	1.05	0.0148	18	769	731	11.7	0.89	0.0021
30	261	239	0.224	1.05	0.0126	19	880	842	13.0	0.86	0.0018
31	283	260	0.247	1.04	0.0116	20	0.64	0.13	0.164	25.4	4.20
32	333	311	0.319	1.03	0.0098	21	1.56	0.30	0.158	9.65	1.67
33	393	368	0.340	1.05	0.0085	22	2.61	0.50	0.160	5.9	1.00
34	588	561	0.502	1.01	0.0055	23	3.40	0.68	0.167	47	0.76
B. Copper pipe, $R=1.30$ cm, $L=332$ cm, hydrogen, using 4-in. oil diffusion pump and Megavac.						24	0.45	1.10	0.172	36.2	5.65
1	4	3	0.093	15.3	1.66	25	1.00	0.22	0.179	16.7	2.54
2	8	6	0.079	6.45	0.83	26	0.60	0.10	0.144	23.4	4.44
3	9	7	0.091	6.52	0.73	27	2.05	0.40	0.170	7.96	1.27
4	11	8	0.089	5.39	0.61	E. Iron pipe, $R=5.12$ cm, $L=457$ cm; air, using N.R.C. oil diffusion pump E-10 and 100 cfm Kinney mechanical pump.					
5	16	12	0.091	3.72	0.41	1	101	37	10.5	1.02	0.0116
6	21	16	0.091	2.83	0.31	2	73	24	8.10	1.11	0.0163
7	17	15	0.093	3.33	0.36	3	60.9	17.4	6.26	1.07	0.0202
8	17	15	0.092	3.32	0.36	4	45.0	11.5	4.57	1.08	0.0283
9	17	15	0.097	3.49	0.36	5	35.0	8.2	3.65	1.13	0.037
10	18	16	0.107	3.62	0.34	6	23.8	4.6	2.58	1.21	0.056
11	25	22	0.103	2.52	0.26	7	17.6	3.1	2.02	1.30	0.077
12	27	24	0.114	2.57	0.23	8	13.0	2.2	1.71	1.51	0.105
13	29	26	0.129	2.69	0.21	9	8.0	1.1	1.19	1.74	0.174
14	40	36	0.143	2.18	0.153	10	5.8	0.8	1.04	2.11	0.242
15	48	43	0.143	1.81	0.138	11	3.99	0.44	0.77	2.36	0.366
16	55	50	0.172	1.88	0.110	12	2.5	0.30	0.70	3.35	0.57
17	61	55	0.162	1.60	0.100	13	1.14	0.11	0.60	6.4	1.27
18	71	65	0.198	1.67	0.085	14	0.52	0.06	0.62	14.2	2.76
19	82	75	0.205	1.50	0.074	15	0.29	0.02	0.57	24	5.0
20	125	116	0.271	1.30	0.048	16	1.80	0.20	0.65	6.5	0.80
21	133	124	0.290	1.30	0.044	17	435	413	50	0.79	0.0019
22	295	284	0.645	1.28	0.020	18	319	289	34	0.73	0.0026
23	327	314	0.620	1.11	0.0181	19	197	164	27	0.99	0.0044
24	406	393	0.773	1.11	0.0145	20	160	118	17.6	0.85	0.0058
25	595	580	1.08	1.06	0.0093	21	95.0	34.0	9.3	0.97	0.0124
26	650	635	1.08	1.05	0.0090						
27	681	665	1.19	1.02	0.0086						
28	758	742	1.27	0.97	0.0077						
29	977	960	1.77	1.05	0.0060						
30	1017	1000	1.80	1.03	0.0058						

TABLE II (Continued).

Run	μ	μ	U	F	$\frac{F}{\mu R(\rho L)^{1/2}}$
F. Iron pipe, $R=10.1$ cm, $L=1040$ cm; air, using N.R.C. oil diffusion pump E-10 and 100 cfm Kinney mechanical pump.					
1	84	72	62	0.79	0.0052
2	80	67	54	0.73	0.0055
3	48	30	34	0.88	0.0104
4	32.5	12	19.9	0.91	0.0184
5	21.9	5.7	13.2	0.96	0.029
6	16.9	4.2	11.2	0.97	0.038
7	11.5	2.7	8.1	1.14	0.057
8	5.6	1.0	4.6	1.40	0.123
9	4.0	0.6	3.6	1.56	0.176
10	2.1	0.3	2.6	2.20	0.34
11	1.35	0.24	2.5	3.1	0.51
12	1.1	0.20	2.3	3.6	0.62
13	0.38	0.12	2.6	10.4	1.62
14	0.25	0.05	2.06	13.8	2.70
15	0.15	0.05	3.5	35	4.05

CORRELATION OF DATA FROM THE LITERATURE

Theoretical equations are often of value in suggesting dimensionless groups to be used as a basis for an empirical correlation, even though the data and theory do not agree quantitatively. In the present instance Eqs. (8) and (14) suggest the probability of a relation between F , X , and f . Figures 1 and 2 show the result of plotting F vs. X for the various data in the literature^{3, 6, 9-11} on flow of air, H_2 , O_2 , N_2 , and CO_2 at low pressures in round glass tubes of small diameter. The correlation for the five gases is excellent, in spite of the 200-fold variation in tube radius. The solid line shown on Fig. 2 represents

$$F = 8.0X$$

corresponding to $f=0.77$. In the range of higher pressures shown in Fig. 1 ($X < 1.0$), the solid line best representing the data corresponds to $f=0.84$.

Figure 2 indicates measured values of F as large as 40,000. This means that the flow at low pressures reaches rates as great as 40,000 times that which would be predicted by Poiseuille's equation, i.e., the effective viscosity of the gas at low pressures may be as small as one forty-thousandth the viscosity at one atmosphere.

³ G. W. Todd, Univ. Durham Phil. Soc. Proc. 6, Part I, 8 (1920).

⁶ W. Klose, Physik. Zeits. 31, 503 (1930).

¹¹ W. Gaede, Ann. d. Physik 41, 289 (1913).

EXPERIMENTAL

Although the correlation of data from the literature is satisfactory, it is evident that extrapolation from glass tubes 7 mm and less in radius to iron pipes several inches in diameter may be considered to be questionable. Accordingly, new data have been obtained for copper and iron pipe $\frac{1}{8}$ to 8 in. nominal diameter. Data were also obtained on $1\frac{1}{2}$ - and 3-in. elbows and a $1\frac{1}{2}$ -in. Kinney high vacuum bellows valve.

The first experiments were carried out using a copper tube 1.30 cm radius and 332 cm between taps to two McLeod gauges. Carefully dried air was admitted through a vacuum reducing valve to a 5-liter vessel from which it passed through an open mercury seal to the test pipe. After a steady state had been reached the mercury seal was closed and the rate of increase of pressure in the 5-liter vessel measured by means of a McLeod gauge. The gas feed rate was calculated from the rate of pressure increase and the volume of the vessel. Pumping was accomplished by means of a 4-inch oil diffusion pump and a Megavac pump in series. A thermocouple gauge

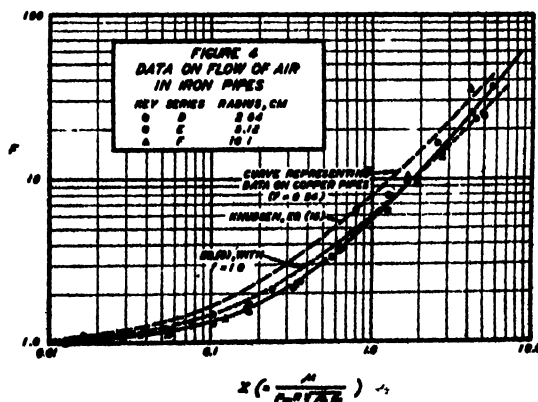
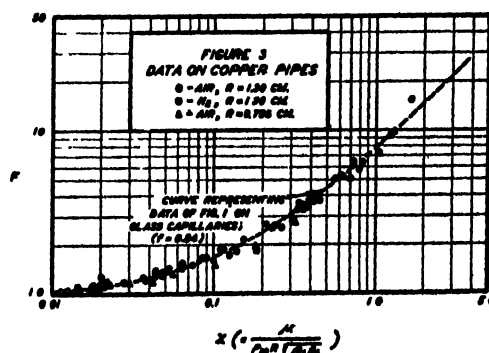


TABLE III. Flow of air in elbows and valves.

Run	p_1	p_2	U	Re	F
A. Eight 1½-in. elbows and 277 cm straight 1½-in. nominal pipe size ($R=2.04$ cm) iron pipe between pressure taps.					
1	4.4	0.6	0.089	0.025	-1.7
2	2.16	0.26	0.074	0.0104	-0.5
3	7.1	1.1	0.100	0.045	-1.9
4	22.1	4.1	0.135	0.179	-0.6
5	49.0	11.0	0.215	0.604	0.6
6	81.0	21.0	0.36	1.58	0
7	120.0	38.0	0.51	3.08	0.5
8	144.5	63.5	0.70	4.2	0
9	150.0	108.0	0.84	2.6	0.1
10	238.0	180.0	1.28	5.5	0.3
11	300.0	230.0	1.50	7.75	0.9
12	410.5	343.5	2.5	12.6	-0.5
13	504.5	429.5	3.3	18.3	-1.0
14	637.5	562.5	4.3	24	-1.1
15	790.0	690.0	4.5	33	0.2
16	980.0	870.0	5.4	44	0.7
17	1016.5	919.5	6.1	44	-0.2
18	706.0	620.0	4.9	31	-1.4
19	1201	1099	7.0	53	0.2
20	1971	1849	9.5	85	1.8
21	3070	2930	12.4	129	4.3
22	3978	3821	14.7	170	2.5
23	3677	3523	15.2	173	4.1
24	5391	5209	19.3	260	6.0
25	7406	7194	22	347	9.0
26	12029	11771	27	520	14.8
27	15753	15447	29	663	20
28	17706	17494	45	705	12
29	34380	34020	52	1380	26
B. Six 3-in. elbows and 700 cm straight 3-inch nominal pipe size ($R=3.9$ cm) iron pipe between pressure taps.					
1	0.75	0.07	0.185	0.0048	0.8
2	1.40	0.10	0.190	0.0095	0.1
3	2.60	0.30	0.21	0.0187	-1.7
4	4.60	0.40	0.24	0.039	-1.2
5	7.95	0.85	0.28	0.077	0
6	13.05	1.35	0.38	0.171	-0.7
7	19.0	3.0	0.47	0.29	0.3
8	29.0	5.0	0.64	0.59	0.7
9	40.0	8.2	0.67	0.82	5
10	48.0	10.0	0.98	1.44	1.4
11	62.0	20.0	1.46	2.36	0.1
12	52.5	11.5	1.17	1.85	0
13	96.0	62.0	2.26	3.0	2.2
14	138.5	109.5	3.3	3.7	3.6
15	178.0	150.0	4.2	4.5	4.3
16	255.0	237.0	9.1	6.3	-1.6
17	430.0	410.0	13.0	10.0	1.0
18	4810	4790	168	129	-0.1
19	8413	8387	215	215	42
20	14020	13980	230	350	15
21	22018	21982	320	440	19
22	48031	47970	440	1070	38
23	69040	68060	540	1600	48

TABLE III (Continued).

Run	p_1	p_2	U	Re	L_e
C. Four 1½-in. Kinney vacuum-tight Bellows valves and 226 cm 1½-in. nominal pipe size ($R=2.04$ cm) iron pipe between pressure taps.					
1	27	5	0.154	0.25	1.8
2	56.75	12.25	0.25	0.82	3.3
3	35.15	6.85	0.187	0.39	1.7
4	95.55	22.05	0.357	1.95	5.5
5	135.0	40.0	0.50	3.5	6.4
6	165.0	67.0	0.65	4.7	5.9
7	210.0	116.0	0.90	6.2	5.9
8	265.0	165.0	1.10	8.1	6.9
9	318.0	220.0	1.33	9.7	7.5
10	384.5	289.5	1.73	12.1	6.7
11	452.0	358.0	2.16	15	6.0
12	370.0	270.0	1.65	12.2	6.3
13	470.0	380	2.8	19	2.0
14	510.0	410	3.0	22	2.4
15	731.5	618.5	3.9	32	4.2
16	2897	2704	9.1	130	18.5
17	4732	4468	11.2	218	29
18	7574	7226	13.5	350	44
19	11830	11370	15.3	520	66
20	15580	15020	16.1	660	85
21	13420	12980	18.1	580	62
22	26590	26010	23	980	106
23	45980	45020	22	1580	203

air flow was measured by an oil displacement device connected upstream from the vacuum reducing valve controlling the flow rate. Air normally passed first through an open stopcock and then through the reducing valve. A large graduated U-tube filled with oil of low vapor pressure was connected to the system between the two valves—the other arm of the U-tube was open to the atmosphere. When the first stopcock was closed air was withdrawn from one arm of the U-tube and the oil level changed at a rate corresponding to the air flow. The actual flow rates were calculated from the observed rate of change of oil level in the U-tube. For very high rates calibrated brass nozzles were employed and air flow rates obtained from observed upstream pressures.

The experimental data on the several straight pipes are summarized in Table II.

RESULTS OF TESTS WITH COPPER AND IRON PIPES

The data of the two series on copper pipes is shown in Fig. 3, where the correction factor F to be applied to Poiseuille's equation (1) is plotted vs. the dimensionless group $\mu/p_m R(\rho_1 g_0)^{1/2}$. The dotted line shown is that from Figs. 1 and 2 for the data from the literature on small glass capillaries. This curve also represents Eq. (8) for

near the pipe inlet served to indicate when steady conditions had been attained. All tests were at approximately 26°C. Both air and hydrogen were used in this equipment.

The second series was carried out in a similar way, with provision of higher pumping rates to handle larger pipes. Low pressures were measured with both McLeod and Alphatron gauges. The

$f=0.84$. Not only is the correlation good, but the data evidently agree well with the previous data on small glass tubes.

Figure 4 shows the data on three sizes of large iron pipes. Again the correlation is good, although the solid line representing the data shows a greater curvature than the lines of constant f , and the points near the middle of the curve correspond to values of f greater than unity. The data fall 0 to 30 percent below the curve for glass and copper tubes. The two smaller iron pipes were not new, but were fairly clean, while the largest pipe was dirty and rough. Since the points for all three pipes fall on a single curve no variation in f with pipe conditions is indicated. The solid curve may be taken as a satisfactory empirical correlation of the data on the three sizes of iron pipe.

Maxwell⁴ introduced the factor f to allow for the fact that some of the molecules striking the wall are diffusely reflected and some are specularly reflected. The physical meaning of f would suggest that it should not vary with flow conditions, and that it should have the same value in the slip-flow region as in molecular flow. Actually the data on glass capillaries are best fitted by taking f as 0.77 for molecular flow ($X > 1.0$) and f as 0.84 for slip-flow ($X < 1.0$). Furthermore, the data on iron pipes correspond to a considerable variation of f through the slip-flow region. It seems probable that these results are caused by faults in the theory, rather than to real variations in f . In the slip-flow region the velocity distribution across the pipe would be expected to become flatter as the pressure decreases and internal friction or viscosity becomes less and less significant, and not to maintain its original parabolic shape superimposed on the increasing velocity of slip at the walls. This effect, which is not taken into account in the derivation for slip-flow, would be expected to be greater in the intermediate region of slip-flow, where both internal and wall friction are important. It is significant that the data deviate most from curves of constant f in this region, and that the deviation is in the right direction.

It will be noted that in Table II there appear several points for which F is less than unity. For the copper pipes there are nine tests with X less than 0.01, with an average F of 1.03. For

the iron pipes there are fifteen such tests, with an average F of 0.88. This indicates that the observed flow rates in streamline flow are lower than would be calculated by Poiseuille's equation, and suggests that the whole curve of Fig. 4 may be somewhat low. For most purposes this error, if applicable to flow at low pressures, will introduce a small factor of safety in using the solid curve of Fig. 4 as an empirical correlation for iron pipes in design calculations.

TABLE IV. Flow of air in short lengths of straight iron pipe.

Run	p_1	p_2	U	X	K
A. 23.4 cm of 1-in. pipe ($R=1.33$ cm).					
1	397	27.0	1.57	0.0145	0.058
2	246	4.0	1.16	0.0246	0.070
3	199.5	2.5	1.05	0.0305	0.077
4	155	2.0	0.90	0.039	0.083
5	110	1.0	0.74	0.055	0.091
6	61.1	0.5	0.54	0.10	0.105
7	32	0	0.36	0.192	0.109
8	16.8	0	0.28	0.36	0.109
9	24	0	0.32	0.257	0.109
10	13.5	0	0.27	0.46	0.114
11	6.15	0.05	0.25	1.0	0.12
12	2.90	0	0.23	2.05	0.109
13	2.20	0	0.23	2.8	0.109
14	1.25	0	0.23	4.9	0.101
15	0.88	0	0.22	7.0	0.089
B. 12.7 cm of 1-in. pipe ($R=1.33$ cm).					
1	334	31.0	1.94	0.017	0.081
2	216.5	3.5	1.49	0.028	0.10
3	148	0	1.29	0.042	0.123
4	119.5	1.5	1.12	0.051	0.129
5	97.5	6.5	1.01	0.066	0.142
6	74	1.0	0.89	0.082	0.150
7	55.5	0.5	0.80	0.11	0.168
8	41.0	0.4	0.74	0.149	0.191
9	32.5	0.5	0.64	0.187	0.190
10	25	0	0.50	0.246	0.168
11	15	0	0.43	0.41	0.173
12	10.6	0.1	0.40	0.58	0.175
13	13	0	0.40	0.47	0.172
14	7.3	0	0.36	0.83	0.168
15	3.1	0	0.38	1.99	0.18
16	1.7	0	0.35	3.42	0.16
17	0.88	0	0.33	7.0	0.13
C. 6.4 cm. of 1-in. pipe ($R=1.33$ cm).					
1	305	51	2.55	0.017	0.111
2	160	4	1.8	0.037	0.159
3	119	2	1.65	0.05	0.188
4	79	1	1.42	0.077	0.23
5	53	1	1.27	0.114	0.27
6	35.5	0.5	1.09	0.171	0.30
7	28	0	0.95	0.22	0.31
8	18	0	0.79	0.34	0.31
9	10.8	0	0.70	0.57	0.31
10	5.0	0	0.80	1.23	0.38
11	3.2	0	0.60	1.92	0.28
12	1.6	0	0.57	3.8	0.26
13	0.58	0	0.62	10.6	0.27
14	0.40	0	0.55	15.4	0.24

COMPARISON WITH THE KNUDSEN EQUATION

The Knudsen semi-empirical equation^{12, 13} may be written

$$Q = \frac{\pi g R^4}{16 \mu L} (p_1^2 - p_2^2) \times \left[1 + 8.52X \left(\frac{X+2.00}{X+2.47} \right) \right] \quad (15)$$

or,

$$F = 1 + 8.52X \left(\frac{X+2.00}{X+2.47} \right). \quad (16)$$

This is found to fit the curves of Figs. 1–3 within five percent or less, and so represents the data on glass and copper tubes essentially as well as the empirical lines drawn to represent the data. The Knudsen equation does not, however, fit the data on iron pipe, as indicated by Fig. 4, which includes a dotted curve representing both the copper and glass tube data and Eq. (15).

RESULTS WITH ELBOWS AND VALVES

Table III gives the data obtained by the same technique, using lines in which were inserted: (a) eight 1½-inch elbows, connected by short (6–8 inch) sections of 1½-inch nominal diameter iron pipe; (b) six 3-inch elbows, connected by 36-inch sections of 3-inch nominal diameter iron pipe; and (c) four 1½-inch Kinney vacuum-tight bellows valves connected by sections of 1½-inch iron pipe. The last two columns in each case give values of the Reynolds number, Re , and the length of straight pipe, L , expressed in pipe diameters, corresponding to the resistance of one valve or elbow. Values of L were obtained by comparing the observed values of S with those calculated from (13) for the actual total length of pipe, using values of F from the solid line of Fig. 4.

The results indicate that neither the valves nor the elbows offer appreciable resistance at low pressures. For Reynolds numbers below 100, L can be taken as unity for the elbows, and may be assigned the nominal value of five for the Kinney valve. As Re increases above 100, L , for

the valves becomes appreciable, approaching the accepted value^{14, 15} of about 30 for turbulent flow ($Re > 2100$). In the case of the valve L , is considerably greater, increasing from 45 at a Re of about 50 to greater than 200 at the critical value of Re . The data for the elbows are erratic in the range of Re from 100 to 2000, for which reliable correlations exist.¹⁶ The data show clearly, however, that the resistance of elbows may be neglected at low pressures, where Re is less than 100.

RESULTS WITH SHORT TUBES

Data were obtained on flow through three short lengths of 1-inch iron pipe connecting two large domes. Air was supplied to one dome through a controlled leak, and the pressures in both domes measured. The data obtained are given in Table IV.

Values of S were calculated from Eq. (13) and the solid line of Fig. 4. In the molecular flow region the experimental data are represented by Eq. (11) and the calculated values of S by Eq. (9), using f of approximately 0.9. Hence K is obtained as

$$\frac{\pi}{2} \left(\frac{2}{f} - 1 \right) \frac{R}{L}$$

times the ratio of experimental to calculated values of S . The values of K obtained for the molecular flow region average 0.11, 0.16, and 0.29 in the three series, which compare well with the values 0.12, 0.20, and 0.31 given by Clausing⁷ for the three values of L/R . The Clausing factor K is not properly applicable outside the molecular flow region, as indicated by the way the tabulated values vary and decrease very considerably at higher pressures.

The combined pressure drop due to the contraction and the flow through the tube may be obtained by adding the pressure drops calculated from Eq. (10) for the entrance and Eq. (13) for flow in the straight pipe. If p_1 represents the pressure before contraction, p_1' the pressure in

¹² S. Dushman, J. Frank. Inst. 211, 689 (1931).

¹³ L. B. Loeb, *The Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1934), second edition, Chapter VII.

¹⁴ J. H. Perry, Editor, *Chemical Engineers' Handbook* (McGraw-Hill Book Company, Inc., New York, 1941), second edition, p. 822.

¹⁵ R. E. Wilson, W. H. McAdams, and M. Seltzer, Ind. Eng. Chem. 14, 105 (1922).

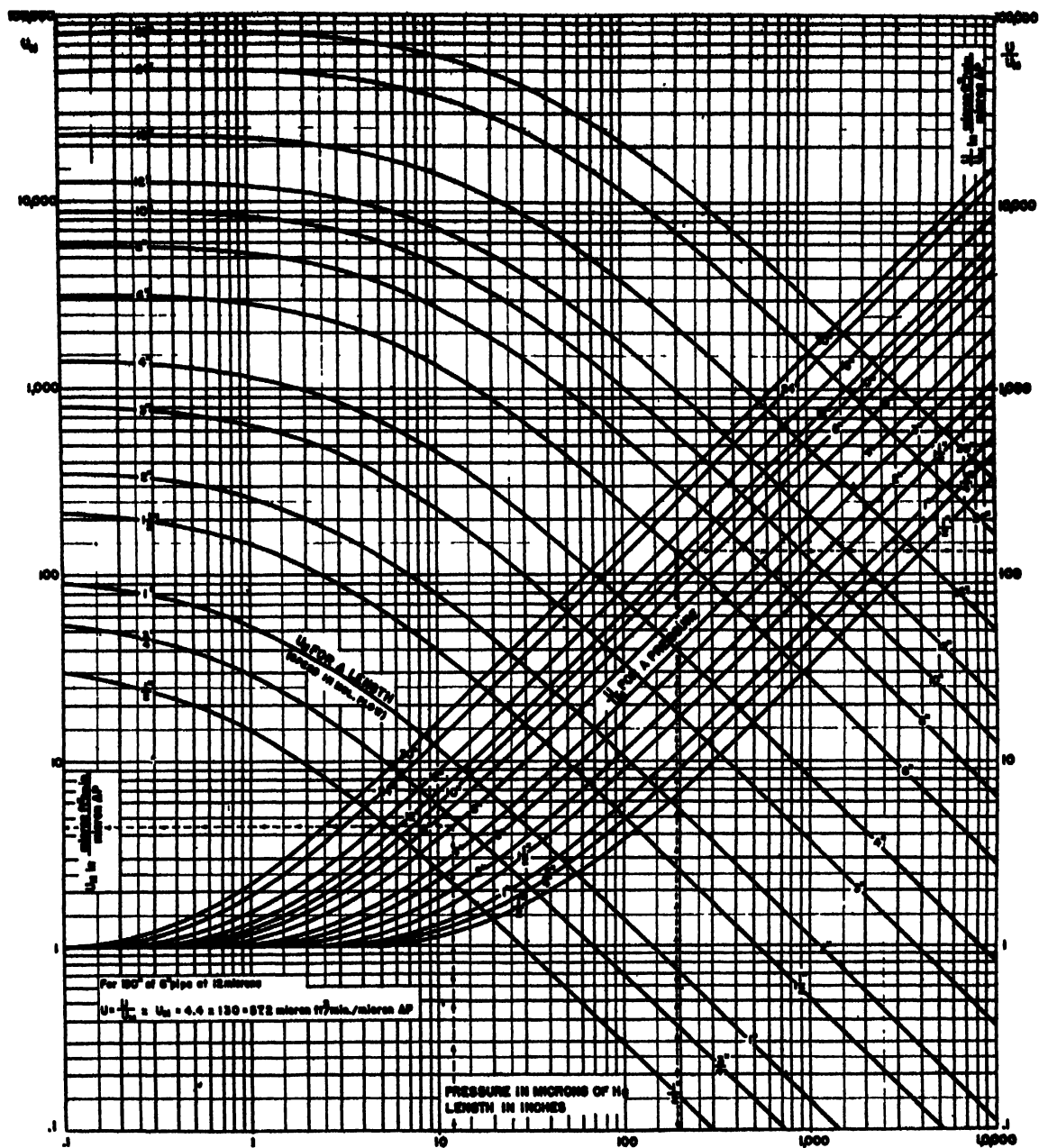


FIG. 5.

the tube at the inlet end, and p_2 the pressure at the downstream end, then

$$\frac{1}{S} = \frac{p_1 - p_2}{Q} = \frac{p_1 - p_1'}{Q} + \frac{p_1' - p_2}{Q}$$

$$= B \left(\frac{2\rho_1}{\mu g_0 R^4} \right)^{\frac{1}{2}} + \frac{8\mu L}{F\pi R^4 p_{m g_0}} \quad (17)$$

In order to compare this equation with the data of Table IV, it was assumed that a correction term B would necessarily be introduced as a multiplying factor before the first term. A value of p_1' was arrived at by trial and error calculation such that $p_1' - p_2$ agreed with the pressure drop calculated from Eq. (13) using $\bar{P}$ obtained using Fig. 4 and $p_m = (p_1' - p_2)/2$. The values of B

obtained in this way varied somewhat due to the fact that the first term of the right side of Eq. (17) was a small difference of two larger numbers, but the average value was about 0.5. It is suggested, therefore, that Eq. (17) be used with a value of B of 0.5 to allow for entrance and exit losses in flow through short tubes. Values of U calculated in this way for the tests of Table IV showed an average deviation of less than 10 percent from the measured values.

As an example, suppose it is desired to calculate S for flow of air at 26°C through a tube 12.7 cm long and 1.33 cm radius when the pressures outside the tube upstream and downstream are 97.5 and 6.5 microns, respectively. Using microns-c.g.s. units, with $\mu = 1860 \times 10^{-7}$ poises, and

$$p_1 = \frac{29}{22,400} \times \frac{273}{299 \times 760,000} = 1.56 \times 10^{-9},$$

then

$$X = \frac{\mu}{p_m R (\rho_1 g_0)^{\frac{1}{2}}} = \frac{1860 \times 10^{-7}}{p_m \times 1.33 (1.33 \times 1.56 \times 10^{-9})^{\frac{1}{2}}} = \frac{3.05}{p_m}.$$

Assume $p_1' = 90$ microns, then $p_m = 48.2$, $X = 0.064$, and from Fig. 4, $F = 1.21$. Substituting in Eq. (17),

$$\begin{aligned} \frac{1}{S} &= 0.5 \left(\frac{2 \times 1.56 \times 10^{-9}}{3.14 \times 1.33 \times 1.33^4} \right)^{\frac{1}{2}} \\ &+ \frac{8 \times 1860 \times 10^{-7} \times 12.7}{1.21 \times 3.14 \times 1.33^4 \times 48.2 \times 1.33} \\ &= 0.77 \times 10^{-5} + 2.47 \times 10^{-5} = 3.24 \times 10^{-5}. \end{aligned}$$

These two terms should be proportional to the two parts of the total pressure drop, so

$$p_1' = 6.5 + \frac{2.47}{3.24} (97.5 - 6.5) = 76 \text{ microns,}$$

which does not check the assumed value of 90. Repeating the calculation until a check is obtained, it is found that $p_1' = 77.5$, $S = 29,600$ micron-cu. cm per sec. per micron, or $U = 1.01$

micron-cu. ft. per sec. per micron. This checks the observed value obtained in Run B-5 (Table IV).

SUMMARY

Data on flow of gases in round pipes at low pressures are well correlated by the equation,

$$S = \frac{Q}{(p_1 - p_2)} = \frac{F \pi R^4 p_m g_0}{8 \mu L}, \quad (13)$$

where F is related empirically to the dimensionless group X by Figs. 1-3 for glass and copper, and Fig. 4 (solid line) for iron pipes. The pressure p_m is the arithmetic mean of the upstream and downstream pressures, p_1 and p_2 . Any consistent set of units may be employed, such as the four given in Table I, and the correlation may be used for streamline flow, slip flow, or molecular flow.

In c.g.s.-micron units the following simplifications may be used for air or water vapor at 26°C.

$$\text{For air, } X = \frac{4.10}{p_m R}; \quad S = 2910 \frac{F R^4 p_m}{L}.$$

$$\text{For water vapor, } X = \frac{2.66}{p_m R}; \quad S = 5500 \frac{F R^4 p_m}{L}.$$

(In units of cu. ft., microns, and minutes, with R and L in inches, the constants become 1.61, 101, 1.05, and 191 in place of 4.10, 2910, 2.66, and 5500, respectively.)

For Reynolds numbers below 100 (most vacuum engineering) the resistance of $1\frac{1}{2}$ - and 3-inch elbows may be neglected; a $1\frac{1}{2}$ -inch Kinney bellows valve has a resistance equivalent to about five pipe diameters.

For flow through short tubes involving both contraction and expansion, the expansion loss is neglected and the flow or pressure drop calculated from Eq. (17), using the method of calculation illustrated by a numerical example in the preceding section.

As an example of the method of calculation, assume that it is desired to determine the "speed" of a standard 2-inch iron pipe with upstream and downstream pressures of 30 and 2.0 microns, respectively, at points 10 feet apart.

The gas is water vapor at 26°C.

$$p_m = \frac{30.0 + 2.0}{2} = 16.0 \text{ microns,}$$

$$R = 2.07 \times 2.54 / 2 = 2.63 \text{ cm,}$$

$$L = 10 \times 30.48 = 304.8 \text{ cm,}$$

$$X = \frac{2.66}{16.0 \times 2.63} = 0.063,$$

and from Fig. 4, $F = 1.41$,

$$S = 5500 \frac{1.21 \times 2.63^4 \times 16}{304.8} \\ = 16,500 \text{ micron-cu. cm/(sec.)(micron),}$$

$$U = 16,500 / 28,400 \\ = 0.58 \text{ micron-cu. ft./(sec.)(micron).}$$

The Reynolds number is $2Q\rho_1/\pi\mu$, and since

$$\rho_1 = \frac{18}{82.07 \times 299 \times 760,000} = 9.65 \times 10^{-10} \text{ g/cu. cm}$$

and $\mu = 950 \times 10^{-7} \text{ g/(sec.)(cm)}$, then

$$Re = \frac{2 \times 16,500(30 - 2) \times 9.65 \times 10^{-10}}{3.14 \times 950 \times 10^{-7}} = 2.99.$$

Since this is well below 100, it may be assumed that the resistance of elbows will be negligible.

Equation (13) and the graphical correlations of F and X may be combined in the form of a single graph or alignment chart for engineering use. Figure 5 is an example of a useful chart prepared in this way, based on a constant f of 0.90. In using the chart, the speed in molecular flow is first obtained from the ordinate of the appropriate diameter curve, using length in inches as abscissa and the curves sloping downward to the right. The ratio of actual speed to speed in molecular flow is then obtained as the ordinate using the arithmetic mean pressure p_m

in microns as abscissa, and one of the family of lines sloping up to the right. The product of the two figures is U , the admittance of the pipe, and may be used in the usual reciprocal speed formula. The curves hold for all combinations of pressure and pipe size except for very short lengths in the streamline region, where the Clausing correction to U_m loses meaning. Fortunately, however, speed evaluations for short pipe at high pressure is seldom critical.

TABLE OF NOMENCLATURE

The units quoted are in the micron-c.g.s. system except where otherwise noted; see the paragraph "Units" regarding use of other systems.

B = correction factor to be applied to the molecular effusion equation to give pressure drop because of contraction.

f = fraction of molecules striking wall which is diffusely reflected.

g_s = factor for converting microns Hg weight to microns Hg mass, = 1.33.

k = coefficient of external friction.

K = Clausing coefficient, in Eq. (11).

L = tube length, cm.

L_e = length of straight pipe giving same pressure drop as valve or elbow diameters.

M = molecular weight of gas.

p_1 = upstream pressure, microns Hg.

p_2 = downstream pressure, microns Hg.

p_m = arithmetic mean pressure = $(p_1 + p_2)/2$.

Q = gas flow rate, micron-cu. cm/second.

R = tube radius, cm.

R_G = gas law constant = 6.24×10^{-7} .

S = rate of flow per unit pressure drop, micron-cu. cm/(sec.)(micron).

T = temperature, °K.

U = rate of gas flow per unit pressure drop = $Q/\Delta p$, micron-cu. ft./(sec.)(micron).

u_0 = velocity of slip at wall, cm/sec.

$U_m = U$ in molecular flow.

$X = \mu/p_m R(\rho_1 g_s)^{1/2}$.

λ = mean free path of gas molecules, cm.

λ_m = mean free path of gas molecules at mean pressure p_m , cm.

ρ_1 = gas density at one micron Hg, g/cu. cm.

μ = gas viscosity at atmospheric pressure, poises.

Absorption of Microwaves by Gases. II

JOHN E. WALTER* AND W. D. HERSHBERGER

RCA Laboratories, Princeton, New Jersey

(Received May 2, 1946)

The absorption coefficients and dielectric constants of sixteen gases have been measured at the two wave-lengths $\lambda=1.24$ cm and $\lambda=3.18$ cm. The gases are H_2S , COS , $(\text{CH}_3)_2\text{O}$, $\text{C}_2\text{H}_6\text{O}$, $\text{C}_2\text{H}_5\text{Cl}$, SO_2 , NH_3 , six halogenated methanes and three amines. Certain improvements in technique are described; these improvements permit detection of absorption coefficients as small as $0.2 \times 10^{-4} \text{ cm}^{-1}$ and measurement of larger coefficients with an accuracy of ± 5 percent. The measured dielectric constants at these wave-lengths are essentially equal to the static values. A quantitative interpretation of the absorption coefficients in terms of the known structure and spectra of the individual molecules is given. The theory indicates that all non-planar molecules which possess a permanent dipole moment should show appreciable absorption in the microwave region.

INTRODUCTION

RESULTS of an investigation on the absorption of microwaves by gases were given in an earlier paper.¹ Measurements on absorption are taken by noting the power loss suffered by a microwave transmitted through a rectangular waveguide operating in the TE_{10} mode. Dielectric constants are measured by comparing the wave-length in the guide when evacuated with the wave-length after the gas under study is admitted to the guide. Measurements are reported at two wave-lengths: 3.18 cm and 1.24 cm. The experimental values of the absorption coefficients α (in cm^{-1}) are given in Table I; the experimental values of the dielectric constants are listed in Table II. The purpose of the present paper is to

TABLE I. Observed absorption coefficients.

	$\alpha \times 10^4$ ($\lambda=3.18$)	$\alpha \times 10^4$ ($\lambda=1.24$)
NH_3	15.5	78.0
CH_3F	7.6	10.0
CH_2Cl_2	5.5	8.0
CH_2Br_2	4.2	6.2
$\text{C}_2\text{H}_5\text{Cl}$	8.7	15.0
CHFCl_2	5.2	10.6
CHF_2Cl	7.2	12.0
CF_2Cl_2	0.3	3.1
H_2S	~ 0.0	~ 0.3
SO_2	1.1	5.0
COS	—	0.8
$(\text{CH}_3)_2\text{O}$	0.7	3.9
$\text{C}_2\text{H}_6\text{O}$	1.0	6.8
CH_3NH_2	2.1	10.5
$(\text{CH}_3)_2\text{NH}$	—	7.0
$\text{C}_2\text{H}_5\text{NH}_2$	3.1	10.2

report improvements in technique, to give the theory underlying the observed absorption, and to present new data and analyses of these data.

TECHNIQUES AND ACCURACY OF MEASUREMENT

One factor which has limited the accuracy with which small absorption coefficients may be measured is the standing wave in the test section of guide. This effect was troublesome in the work reported earlier but has now been eliminated by placing a wedge of lossy solid material in this section near the input window. The loss introduced by the wedge is 10 db and, since it is not inserted until the standing wave has been minimized by the usual tuning adjustments, no measurable standing wave now exists when measurements are taken. The procedure for tuning is first to fill the 30-foot test section with ammonia which effectively isolates the receiver

TABLE II. Dielectric constants.

	$\epsilon_0 \times 10^6$ (calculated)	$\epsilon \times 10^6$ ($\lambda=3.18$)	$\epsilon \times 10^6$ ($\lambda=1.24$)
NH_3	6.3	5.3	5.5
CH_3F	9.2	8.1	8.4
CH_2Cl_2	10.3	9.4	9.9
CH_2Br_2	9.7	9.5	10.0
$\text{C}_2\text{H}_5\text{Cl}$	12.7	11.1	11.6
CHFCl_2	6.2	4.9	5.2
CHF_2Cl	6.7	5.2	5.7
CF_2Cl_2	3.1	2.9	3.1
H_2S	3.6	3.0	3.5
SO_2	7.9	7.5	8.6
COS	2.6	—	2.6
$(\text{CH}_3)_2\text{O}$	6.2	5.5	6.1
$\text{C}_2\text{H}_6\text{O}$	10.4	10.2	10.7
CH_3NH_2	5.3	4.8	5.3
$(\text{CH}_3)_2\text{NH}$	4.4	—	4.0
$\text{C}_2\text{H}_5\text{NH}_2$	6.0	5.3	4.5

* Present Address: Palmer Physical Laboratory, Princeton University, Princeton, New Jersey.

† W. D. Hershberger, J. App. Phys. 17, 495 (1946).

from the generator and then adjust the tuner which is adjacent to the input window for a minimum standing wave. Next, the ammonia is removed and the tuner adjacent to the output window is adjusted for a minimum standing wave. As a result of this adjustment receiver power very nearly attains its maximum value. Finally, the 10-db wedge is placed inside of the test section of guide and the operation of the system is relatively insensitive to changes in the frequency of the generator. The result is that an absorption coefficient as small as 0.2×10^{-4} neper per cm may be measured, but of course the accuracy in this range is low. Thus if the received power falls to 98 percent of its initial value on introducing a gas into a 30-foot section of guide the absorption coefficient is 0.18×10^{-4} neper per cm, while if the power drops to 97 percent of its initial value, the coefficient is 0.27×10^{-4} neper per cm. When a gas with large absorption is used, the fact that absorption vanishes as the pressure is reduced is of great aid in following the course of a pressure vs. absorption curve.

When the absorption coefficient is of the order of 10^{-3} neper per cm or 40 times the minimum measurable absorption, the consistency of the results obtained is such that the accuracy of the absorption coefficients reported is estimated to be somewhat better than 5 percent. Approximately the same accuracy is realized in the measurement of δ , which is the amount by which the dielectric constant exceeds unity.

Figure 1 represents the experimental points obtained in six absorption runs with methyl chloride. These readings lie in the range of pressure and absorption for which accuracy is high, using the 30-foot section of guide.

GENERAL THEORY

The essential parameters which characterize an absorption line are ν_0 , the frequency of the center of the line, and the half-width $\Delta\nu$, which is the difference between ν_0 and the frequency at which the absorption has fallen to one-half of the absorption at the center of the line. In the microwave region the observed broadening of an absorption line is to be attributed to collision broadening; Doppler broadening and natural line breadth are negligible at these frequencies. Since at one atmosphere $\Delta\nu$ is found to be of the same

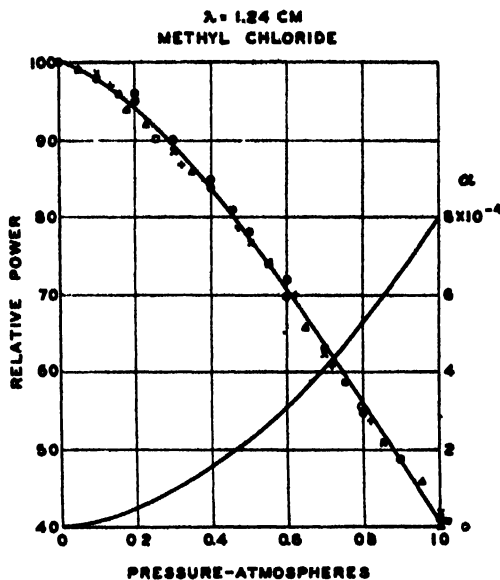


FIG. 1. Absorption of methyl chloride at $\lambda = 1.24$ cm.

order of magnitude as ν_0 if ν_0 is in the microwave region, the exact shape of the absorption line becomes a matter of considerable interest. This problem of the shape of collision-broadened lines has been re-examined by Van Vleck and Weisskopf,² who have shown that the classical formula of Lorentz requires revision. The fundamental difference between the assumptions made in the two theories, in applying them to absorption by molecules possessing a permanent dipole moment, is the following: Lorentz assumed that after each collision all orientations of the dipole with respect to the direction of the electric field are equally probable; Van Vleck and Weisskopf assumed that the probability of a given orientation is given by the corresponding Boltzmann factor. The absorption coefficient α_ν for electromagnetic radiation of frequency ν is calculated by the latter authors to be

$$\alpha_\nu = \frac{8\pi^3 \nu N}{3hc} \cdot \frac{\sum_j \sum_i |\mu_{ij}|^2 f(\nu_{ij}, \nu)}{\sum_j e^{-W_j/kT}} e^{-W_i/kT}, \quad (1)$$

$$f(\nu_{ij}, \nu) = \frac{1}{\pi} \frac{\nu}{\nu_{ij}} \left\{ \frac{\Delta\nu}{(\nu_{ij} - \nu)^2 + \Delta\nu^2} + \frac{\Delta\nu}{(\nu_{ij} + \nu)^2 + \Delta\nu^2} \right\}, \quad (2)$$

² J. H. Van Vleck and V. F. Weisskopf, *Rev. Mod. Phys.* 17, 227 (1945).

where

N = number of molecules per cc,
 μ_{ij} = electric dipole moment matrix element for the two states i and j which have a frequency separation ν_{ij} ,
 W_j = energy of j th state, and
 $\Delta\nu$ = half-width of a line = $1/2\pi\tau$, where τ is the mean time between molecular collisions.

The other symbols have their usual significance. The summations are to be extended over all the states of the molecule. Equations (1) and (2) differ from the Lorentz result by the inclusion of the factor ν/ν_{ij} in (2) and by the change in the sign of the second term in (2). For a molecule with large moments of inertia and a permanent dipole moment μ_0 , these equations reduce to the familiar Debye formula

$$\epsilon'' = \frac{\alpha\epsilon}{2\pi\nu} = \frac{4\pi N\mu_0^2}{3kT} \frac{\nu\Delta\nu}{\nu^2 + \Delta\nu^2} \quad (3)$$

for the complex part of the dielectric constant.

For our purposes it is convenient to specialize Eqs. (1) and (2) so that they apply directly to the ammonia problem. NH_3 is a symmetrical top molecule with rotational energy levels specified by the quantum numbers J and K .⁵ Each rota-

tional level is split into a close doublet as a result of the tunneling effect; the two components of the doublet are specified by the designation $+$ and $-$. The selection rules for rotational transitions are $\Delta J = 0, \pm 1$; $\Delta K = 0$; $+\leftrightarrow -$. The "inversion spectrum" frequency corresponding to $\Delta J = 0, \Delta K = 0$; $+\leftrightarrow -$ is about 0.8 cm^{-1} while the lowest frequency in the rotation spectrum corresponding to $\Delta J = \pm 1$ is about 20 cm^{-1} . For $\nu < 1.0 \text{ cm}^{-1}$ only the absorption associated with the inversion spectrum need be considered, and Eq. (1) can be reduced to⁵

$$\alpha_\nu = \frac{8\pi^2\nu N}{6ckT} \frac{\sum_j \sum_i |\mu_{ij}|^2 f(\nu_{ij}, \nu) \nu_{ij} e^{-W_j/kT}}{\sum_j e^{-W_j/kT}}, \quad (4)$$

where j and i now refer to the two components of a given doublet. The reduction follows from the relation $h\nu_{ij} \ll kT$. Aside from the negligible difference in the Boltzmann factor, we obtain the same contribution to the sum in the numerator from the transition $j \rightarrow i$ as from the transition $i \rightarrow j$. If $W(J, K)$ is the mean energy of a given doublet, and $g(J, K)$ is the total weight of the rotational level with quantum numbers J and K , Eq. (4) becomes

$$\alpha_\nu = \frac{4\pi^2 N \nu^3}{3ckT} \frac{\sum_{J,K} g(J, K) |\mu(J, K)|^2 f'[\nu(J, K), \nu] e^{-W(J, K)/kT}}{\sum_{J,K} g(J, K) e^{-W(J, K)/kT}}, \quad (5)$$

where

$$f'[\nu(J, K), \nu] = \frac{\Delta\nu}{[\nu(J, K) - \nu]^2 + \Delta\nu^2} + \frac{\Delta\nu}{[\nu(J, K) + \nu]^2 + \Delta\nu^2}, \quad (6)$$

and where the sums over the rotational levels in the excited vibrational states have been omitted. $\mu(J, K)$ is the electric moment matrix element for the transition $\Delta J = 0, \Delta K = 0, +\leftrightarrow -$ and is equal to⁴

$$|\mu(J, K)|^2 = \frac{K^2}{J(J+1)} \mu_0^2, \quad (7)$$

where μ_0 is the permanent dipole moment of the molecule. For an accidentally symmetrical top the weights $g(J, K)$ are $(2J+1)$ for $K=0$, and

$2(2J+1)$ for $K \neq 0$. For NH_3 or CH_3Cl the presence of the threefold symmetry axis and the nuclear spin of the hydrogen atoms introduce the additional factor 4 for K divisible by 3 and the factor 2 for K not divisible by 3.⁵ Since both numerator and denominator of (5) involve the same set of weights, only the relative values are important. Factors such as the spin of N or Cl, which would change the absolute weights, will not affect the relative weights. In the sums we can therefore use the weights

$$g(J, K) = (2J+1) \text{ times } 4, 4, 4, 8, 4, 4, 8, 4, \dots$$

⁴ G. Herzberg, *Infra-Red and Raman Spectra* (D. Van Nostrand Company, New York, 1945), pp. 26-34.
⁵ Reference 3, p. 422.

⁵ Reference 3, p. 28.

for

$$K = 0, 1, 2, 3, 4, 5, 6, 7 \dots \quad (8)$$

If these weights are used, the sum in the denominator, which we denote by Q , becomes (in the high temperature approximation)

$$Q = 8/3(Q_{\text{class}}) = 8/3(1.027)[T^2/B^2A]^{\frac{1}{2}}, \quad (9)$$

where Q_{class} is the classical rotational partition function for a symmetrical top with rotational constants B and A , and T is the absolute temperature. The energy levels, in wave numbers, are

$$W(J, K) = BJ(J+1) + (A-B)K^2. \quad (10)$$

If $\nu(J, K) = \nu_0$ independently of J and K , the absorption coefficient is given by

$$\alpha_\nu = \beta N \nu^2 \left[\frac{\Delta \nu}{(\nu_0 - \nu)^2 + \Delta \nu^2} + \frac{\Delta \nu}{(\nu_0 + \nu)^2 + \Delta \nu^2} \right], \quad (11)$$

where β is a constant which can be calculated from the sums in (5). We also note that

$$\int_{-\infty}^{\infty} f'[\nu(J, K), \nu] d\nu = \pi, \quad (12)$$

independently of $\nu(J, K)$ and $\Delta \nu$. Therefore, from (12) and (5), we have, as a measure of the total absorption arising from this inversion spectrum

$$\int_0^{\infty} \frac{\alpha_\nu d\nu}{\nu^2} = \frac{4\pi^2 N \mu_0^2}{3ckT Q} \sum_{J,K} g(J, K) \frac{K^2}{J(J+1)} e^{-W(J,K)/kT} \quad (13)$$

independently of any assumptions as to the detailed positions or breadths of the individual lines.

Since our measurements are limited to two frequencies, they do not give any direct information on the half-widths $\Delta \nu$ of the absorption lines. From the kinetic theory of gases and the relation $\Delta \nu = 1/2\pi\tau$ one would expect to find $\Delta \nu \sim 0.05 \text{ cm}^{-1}$ or less at room temperature and atmospheric pressure. Adel and Barker⁶ find $\Delta \nu = 0.067 \text{ cm}^{-1}$ for zero slit width for the lines in N_2O , which is a linear molecule with a very small dipole moment. For molecules with a large dipole moment much larger values of $\Delta \nu$ have been reported; for ex-

ample, Cornell⁷ reports $\Delta \nu \sim 0.3$ for H_2O and larger values for NH_3 and HCN (not corrected for slit width). These larger values are to be expected in dipole molecules, since a "collision," as the word is applied to line broadening, means that the molecules have come sufficiently close to cause a perturbation of the energy levels, not that they have collided in a billiard-sense. The "optical" collision diameters as computed from observed values of $\Delta \nu$ are usually several times larger than the kinetic theory diameters.

AMMONIA

Although the absorption of microwaves by NH_3 has been known for a long time and the general features understood,⁸ the details of the structure of the line are not clear. Let us assume for the moment that the $\nu(J, K)$'s are all essentially equal, so that the absorption is given by (11). Since $\Delta \nu$ is inversely proportional to the mean time τ between collisions, and since τ is inversely proportional to the pressure P , $\Delta \nu \sim P$. Also, $N \sim P$. For the non-resonant case, with ν_0 very different from ν , Eq. (11) shows that $\alpha_\nu \sim P^2$ as long as $(\nu_0 - \nu) \gg \Delta \nu$. For the resonant case, with $\nu_0 \cong \nu$, α_ν should rise very rapidly with increasing pressure and become relatively independent of P when $\Delta \nu \gg (\nu_0 - \nu)$. These conclusions are in qualitative agreement with the observed behavior of the pressure dependence of the NH_3 absorption (Fig. 2); the curve for $\nu = 0.81 \text{ cm}^{-1}$ ($\lambda = 1.24 \text{ cm}$) is of the type corresponding to $\nu \cong \nu_0$, that for $\nu = 0.32 \text{ cm}^{-1}$ ($\lambda = 3.18 \text{ cm}$) is of the type corresponding to $\nu \ll \nu_0$. The situation is less favorable if looked at from a quantitative viewpoint. The NH_3 curve for $\nu = 0.81$ rises linearly with pressure above 0.1 atmos., and this rise continues to at least 2.5 atmos. The details of the curve cannot be accounted for by Eq. (11) if $\Delta \nu$ is assumed to be proportional to P . Any other variation of $\Delta \nu$ with P would have no theoretical justification at present. If we assume $\nu_0 \cong 0.80 \text{ cm}^{-1}$, α_ν should decrease with frequency more rapidly than ν^2 in the region $\nu < 0.80 \text{ cm}^{-1}$. This is in serious disagreement with the observed decrease. Actually, the $\nu(J, K)$'s are not identical.

⁷ S. D. Cornell, Phys. Rev. **51**, 739 (1937).

⁸ C. E. Cleeton and N. H. Williams, Phys. Rev. **45**, 234 (1934). A complete discussion of the ammonia spectrum and references to the original literature may be found in reference 3, pp. 221, 257, 416.

⁶ A. Adel and E. F. Barker, Rev. Mod. Phys. **16**, 236 (1944).

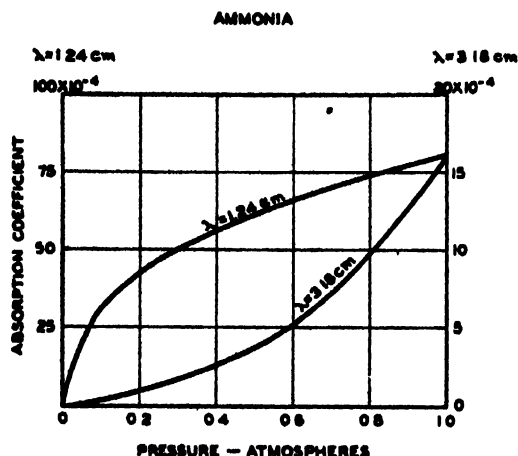


FIG. 2. Absorption of ammonia employing 5-foot guide

The centrifugal distortion of the molecule caused by rotation produces a change in the effective barrier height and thus a change in the doublet separation. This dependence of the doublet separation on J and K has been calculated by Sheng, Barker, and Dennison,⁹ who obtain the result

$$\nu(J, K) - \nu_0 = -0.0011(J^2 + J) + 0.0016K^2, \quad (14)$$

where ν_0 is the doublet separation in the non-rotating molecule. If the Boltzmann distribution is taken into account, one finds that there are lines of appreciable intensity spread over a region about 0.06 cm^{-1} wide (at zero pressure). If the ammonia inversion spectrum consists of a series of lines starting at $\nu \sim 0.80 \text{ cm}^{-1}$ and extending to considerably lower frequencies, the observed features of the absorption curve are more readily understood. This distribution of frequencies would increase the absorption at $\nu = 0.32 \text{ cm}^{-1}$ relative to that at $\nu = 0.81 \text{ cm}^{-1}$ above that expected for a single line near $\nu = 0.80 \text{ cm}^{-1}$, and would also produce a pressure variation similar to that observed at $\nu = 0.81 \text{ cm}^{-1}$. Such a distribution would also explain the apparent discrepancy between the doublet separation as obtained from the pressure curves ($\nu_0 \cong 0.80 \text{ cm}^{-1}$) and from the far infra-red spectra ($\nu_0 \cong 0.66 \text{ cm}^{-1}$) as observed by Wright and Randall.¹⁰ It would seem, however, that a spread greater than that

given by Eq. (14) would be required. Van Vleck and Weisskopf³ have pointed out that a variation of $\Delta\nu$ with J and K would also tend to increase the absorption at the edge of the line relative to the absorption at the center. It appears that a quantitative description of the NH_3 absorption must be postponed until an investigation at low pressures reveals further information about the $\nu(J, K)$'s.

Application of Eq. (13) to NH_3 yields

$$\int_0^\infty \frac{\alpha d\nu}{\nu^2} = 74 \times 10^{-14} \text{ (theoretical)}. \quad (15)$$

(The NH_3 constants are $A = 6.31$, $B = 9.94$, $\mu_0 = 1.49 \times 10^{-18}$. The ratio Σ/Q , where Σ represents the sum in (13), is equal to 0.39 for NH_3 at $T = 300^\circ\text{K}$.) From the data of Cleeton and Williams⁸

$$\int_0^\infty \frac{\alpha d\nu}{\nu^2} \cong 35 \times 10^{-14} \text{ (experimental)}. \quad (16)$$

The origin of this discrepancy is not clear. Comparisons of this type often lead to even more serious disagreements,¹¹ but these are usually attributed to errors arising from finite slit width. Foley and Randall¹² have made a similar comparison on the line $J = 14 \rightarrow J = 15$ of NH_3 at a pres-

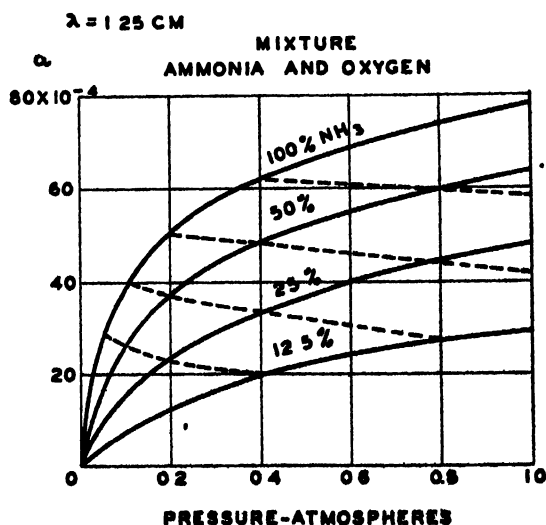


FIG. 3. Absorption of mixtures of ammonia and oxygen.

⁹ H. Y. Sheng, E. F. Barker, and D. M. Dennison, *Phys. Rev.* **69**, 786 (1941).

¹⁰ N. Wright and H. M. Randall, *Phys. Rev.* **44**, 391 (1933).

¹¹ C. Shaefer and F. Matossi, *Das Ultrarot Spectrum* (Verlagbuchhandlung, Julius Springer, Berlin, 1930).

¹² H. M. Foley and H. M. Randall, *Phys. Rev.* **59**, 171 (1941).

sure of 20 cm and have found the experimental value to be only about 20 percent lower than the theoretical value. It would be of interest to see whether or not the above discrepancy persists at low pressures.

Figure 3 shows the effect of the admixture of oxygen on the ammonia absorption. Almost identical curves are obtained from $\text{NH}_3\text{--CH}_4$ and $\text{NH}_3\text{--H}_2$ mixtures. The dotted curves connect points of equal NH_3 partial pressure; if these lines were horizontal it would mean that the NH_3 absorption is a function of the NH_3 partial pressure only, since the absorption of O_2 is negligible. The addition of O_2 to NH_3 changes the line breadth $\Delta\nu$ in the latter molecule and so effects the absorption. We can make a rough estimate of the relative collision cross sections for $\text{NH}_3\text{--O}_2$ and $\text{NH}_3\text{--NH}_3$ collisions as follows. If the pressure of NH_3 is greater than 0.1 atmos., the absorption is, according to Eq. (11), essentially proportional to $1/\Delta\nu$ if the concentration of NH_3 is constant. If we follow the dotted line which starts at $\alpha \times 10^4 = 40$, we find that $\Delta\nu$ increases by a factor $(1.00 + 0.48)$ as the total pressure is increased from 0.1 atmos. to 0.8 atmos. by the addition of oxygen. A corresponding increase in pressure caused by the addition of NH_3 should increase $\Delta\nu$ by a factor $(1.00 + 7.00)$. The relative cross sections for the two-collision process are thus in the ratio 14.5:1, and the relative collision diameters about in the ratio 4:1. Since optical collision diameters are considerably larger than kinetic theory diameters, we see that, roughly at least, the $\text{NH}_3\text{--O}_2$ collision diameter is close to the kinetic theory value.

METHYL HALIDES

The methyl halides CH_3F , CH_3Cl , CH_3Br all show the type of variation $\alpha_r \sim P^2$ which is characteristic of absorption at a frequency far removed from the resonant frequency (Fig. 4). In this figure the curves for CH_3Cl are not shown. For this gas, at a given wave-length, the curve absorption coefficient *vs.* pressure lies between the corresponding curves for CH_3F and CH_3Br . The absolute magnitude of the absorption, the variation of α , with ν , and the similarity of the absorption in the three gases are all inconsistent with the assumption that the observed absorption arises from rotational transitions. Comparing the

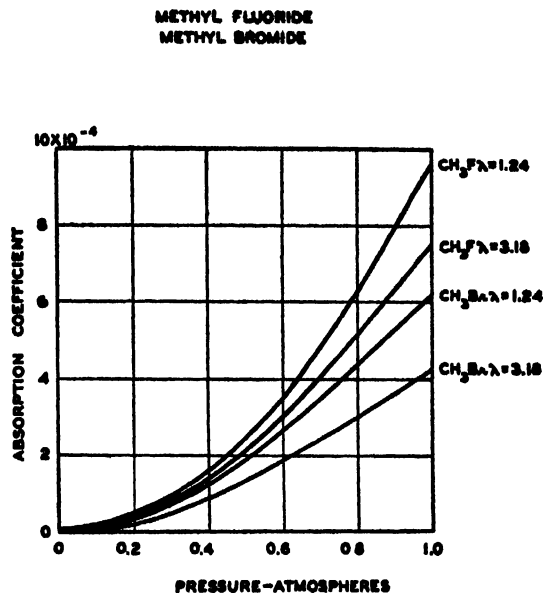


FIG. 4. Absorption of methyl fluoride and methyl bromide.

observed absorption at $\nu = 0.81 \text{ cm}^{-1}$ and $\nu = 0.32 \text{ cm}^{-1}$ with Eq. (11), we conclude immediately that ν_0 is very small.

From a formal viewpoint, the energy levels of CH_3X are exactly the same as those of NH_3 , and the selection rules are identical. Each rotational level characterized by the quantum numbers J and K is double, the two components being designated by the symbols $+$ and $-$. The separation of this doublet depends on the height of the potential barrier which must be overcome if the molecule is to be turned into its mirror image. In NH_3 this barrier (for the passage of N through the plane of the H's) is low and the corresponding doublet separation is large. The potential barrier preventing inversion in CH_3X (perhaps most easily visualized as an internal rotation of a CH_3 group relative to the rest of the molecule) would be expected, on the basis of ordinary chemical concepts, to be much higher than the barrier for inversion in NH_3 , and the doublet separation would be expected to be much smaller. From (11), we see that the absorption becomes

$$\alpha_r = \beta' \nu^2 \Delta\nu / (\nu^2 + \Delta\nu^2), \quad (17)$$

for $\nu \gg \nu_0$, which is independent of the exact value of ν_0 . The experimental results can be fitted to an equation of this form if we assume $\Delta\nu$ proportional to the pressure. The selection rules, statistical weights, and matrix elements are the

same as in the ammonia problem. β' is then given by

$$\beta' = \frac{8\pi^3 N \mu_0^2}{3kT} \sum_{J,K} g(J,K) \times \frac{K^2}{J(J+1)} \cdot e^{-W(J,K)/kT}, \quad (18)$$

if, in (17), the ν 's and $\Delta\nu$'s are in cm^{-1} .

The experimental values of α , were corrected for the contribution arising from rotational absorption; this contribution was estimated from the theory to be 0.00004 for CH_3Cl and 0.00005 for CH_3Br at $\nu = 0.81 \text{ cm}^{-1}$ and negligible in other cases. The experimental values of β' and $\Delta\nu$ were computed from the corrected values of α , for $\nu = 0.81 \text{ cm}^{-1}$ and 0.32 cm^{-1} and $P = 1.0$ atmos. The variation of α , with P is reproduced satisfactorily by Eq. (17) with these values of the constants. The experimental values of $\Delta\nu$ are of the correct order of magnitude. The theoretical values of β' are 30–40 percent higher than the experimental values; this agreement may be regarded as sufficiently exact to verify the correctness of the assumed theory. The results are compared in Table III; it is to be noted that the variation of β' is not identical with the variation in μ_0 since the sums in (18) are appreciably different for the different molecules.

Since we have assumed the doublet separation to be negligibly small, and the frequency associated with the absorption to be essentially zero, we should arrive at the same result if we simply consider the molecules to be rigid rotators, characterized by the quantum numbers J and K . Now a non-rotating dipole molecule can absorb energy from an electromagnetic field because of the effect of collisions; the absorption is given by the Debye formula (3), which can be derived from classical theory. The mechanism can be visualized as the following: After each collision the dipole has a tendency to be oriented parallel to the field; between collisions the dipole is stationary. If the field is varying with time, the

net effect of this mechanism is the production of a polarization which lags behind the field. The polarization current then has a component parallel to the field, resulting in an absorption of energy. The dipole moment vector of a symmetrical top molecule such as CH_3Cl , in the state characterized by J and K , can be resolved into a component parallel to the direction of the total angular momentum vector and a component perpendicular to this direction. The former component is stationary in space, and hence gives a full contribution to the Debye absorption; the latter component has a time average equal to zero and hence gives no Debye absorption. The square of the component of the permanent dipole moment μ_0 parallel to the total angular momentum vector is, since the dipole moment vector is along the symmetry axis, equal to $M_K^2 \mu_0^2 / M^2$ where M_K is the angular momentum about the symmetry axis of the molecule and M^2 is the total angular momentum. From quantum mechanics $M_K^2 = K^2 \hbar^2 / 4\pi^2$, $M^2 = J(J+1) \hbar^2 / 4\pi^2$; the square of this component is thus equal to $K^2 \mu_0^2 / J(J+1)$. If, in Eq. (3), we replace μ_0^2 by this last expression, and sum over all values of J and K weighted by the proper statistical factors, we arrive exactly at Eqs. (17) and (18). In other words, the quantum-mechanical theory and the essentially classical theory of Debye lead to identical results. This is no accident, since the Debye theory is the limiting form of the general theory as the resonant frequency approaches zero.³ It should be emphasized that the Debye theory cannot be applied in the form (3), the essential modification being that μ_0^2 in (3) must be replaced by the square of that component of μ_0 which is stationary in space. Although the two viewpoints concerning the CH_3X absorption are formally equivalent, the Debye picture is perhaps more in agreement with our conventional ideas.

COS

From the known moment of inertia of the linear molecule COS, namely, $1.37 \times 10^{-28} \text{ g-cm}^2$, the wave-length for the transition $J=1 \leftrightarrow J=2$ is very nearly 1.24 cm, the operating wave-length. The wave-length is given by

$$\lambda_0 = \frac{4\pi^2 c I}{h(J+1)}. \quad (19)$$

TABLE III.

	$\Delta\nu$ (exp)	β' (exp)	β' (theor.)	$\mu_0 \times 10^{18}$
CH_3F	0.20	0.0053	0.0067	1.81
CH_3Cl	0.22	0.0037	0.0054	1.87
CH_3Br	0.21	0.0029	0.0038	1.78

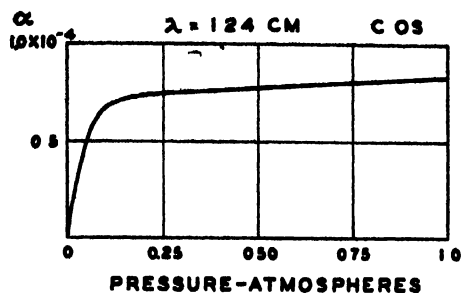


FIG. 5. Absorption of carbonyl sulphide.

The computed absorption is given by

$$\alpha_0 = \frac{2\mu_0^2 \nu_0^2 h^2}{3cI(kT)^2} \times \frac{N}{\Delta\nu}, \quad (20)$$

where μ_0 for COS is 0.65 Debye units, ν_0 is 2.4×10^{10} cycles per second, and N is the number of molecules per cc. At one atmosphere, we obtain the observed value for α_0 , namely 0.8×10^{-4} neper 1 cm, when $\Delta\nu$ is taken to be about 3×10^9 or 0.10 cm^{-1} . The form of Eq. (20) shows that if both N and $\Delta\nu$ vary linearly with pressure, α_0 is independent of pressure. The curve in Fig. 5, showing the absorption of COS, is of the resonant type as expected. In view of the above mentioned result $\Delta\nu = 0.067 \text{ cm}^{-1}$ on the rather similar molecule N_2O , this seems to be a very reasonable value, and indicates a verification of the theory.

H_2S

A survey of the known rotational levels of H_2S^{18} reveals no energy separations in the region $0 < \nu < 3.0 \text{ cm}^{-1}$ which correspond to an allowed transition. The microwave absorption in H_2S should, therefore, be very small, as is observed. The small absorption at $\nu = 0.81 \text{ cm}^{-1}$ ($\alpha_0 \times 10^4 \cong 0.3$) is probably the cumulative effect of the numerous lines at higher frequencies.

SO_2

The rotational levels of SO_2 are not known experimentally, so a detailed calculation is not possible. The levels can be computed if the molecule is considered to be a rigid unsymmetrical top. It is found that there are, up to $J=12$, about 10 lines in the frequency range $0 < \nu < 1.50 \text{ cm}^{-1}$ which correspond to allowed transitions. At $J=12$ the Boltzmann factor is about 0.80, so there will undoubtedly be many strong lines in this frequency range arising from

transitions involving levels higher than $J=12$. It is, therefore, reasonable to assume that the observed absorption in SO_2 could be accounted for on the basis of ordinary rotational transitions, indeed, no other mechanism seems possible.

OTHER MOLECULES

The other molecules listed in Table I are non-planar unsymmetrical tops. As we proceed from a symmetrical top to an unsymmetrical top, we find that each K level splits into two levels, and a transition between these two levels is generally allowed. Transitions of this type, corresponding to $\Delta J=0, \Delta K=0$ in the symmetrical top, will produce a series of lines from $\nu \cong 0$ upward. Also, certain transitions corresponding to $\Delta K \neq 0$ are allowed in unsymmetrical tops. The moments of inertia of these molecules are large; therefore, transitions corresponding to $\Delta J = \pm 1$ will involve low frequencies. The net result is that we have a very complicated rotational spectrum, containing many lines with frequencies from $\nu \sim 0$ upward, and thickly scattered throughout the microwave region. Any exact calculations are almost impossible; however, we should expect the magnitude of the absorption to be about the same as for the methyl halides. The pressure curves often show a behavior intermediate between the resonant and non-resonant types, which is accounted for by the distribution of lines described above. Figure 6 shows such curves for ethylamine.

In general, one would expect all heavy non-planar unsymmetrical top molecules which have a large dipole moment to show an appreciable absorption in the microwave region. Since the selection rules are more stringent for planar and linear molecules, each molecule must be considered individually. Non-planar symmetrical top molecules should be analogous to the methyl halides.

DIELECTRIC CONSTANTS

Van Vleck and Weisskopf³ have given the appropriate formula for the computation of that part of δ which arises from the permanent dipole moment μ_0 . To this must be added the contribution arising from the electronic polarization P_A . For zero frequency, δ is given by

$$\delta_0 = \frac{3(P_E + P_A)}{V} + \frac{4\pi N\mu_0^2}{3kT}, \quad (21)$$

¹⁸ P. C. Cross, *Phys. Rev.* 47, 7 (1935).

where V is the molar volume.¹⁴ δ should not differ very much from δ_0 in the microwave region, since this is on the low frequency side of the region of large absorption. This is illustrated by the comparison of calculated static values and observed values in Table II. The necessary data were taken from Smyth's text,¹⁴ except for the substituted methanes, for which the data were taken from later papers by Smyth and his co-workers.¹⁵

For the methyl halides, the contribution to δ arising from the frequencies near zero (the ones to which we attributed the absorption) is given by

$$\delta_r' = \frac{4\pi N\mu_0^2}{3kT} \frac{\sum \Delta\nu^2}{Q \nu^2 + \Delta\nu^2}, \quad (22)$$

where $\sum$ is the sum in the numerator of (18). For the complex part of the dielectric constant we have

$$\epsilon_r'' = \frac{\alpha_r c}{2\pi\nu} = \frac{4\pi N\mu_0^2}{3kT} \frac{\sum \nu\Delta\nu}{Q \nu^2 + \Delta\nu^2}. \quad (23)$$

If we assume that the contribution to δ arising from the rotational frequencies is equal to the static value of this contribution, the above equations lead to

$$\delta_r = \delta_0 - \epsilon_r'' \frac{\nu}{\Delta\nu} = \delta_0 - \frac{\alpha_r}{2\pi\Delta\nu}. \quad (24)$$

Insertion of the observed values of α_r for CH_3Cl gives $\delta \times 10^3 = 9.9$ at $\lambda = 3.18$ and $\delta \times 10^3 = 9.7$ at $\lambda = 1.24$, as compared with the observed values 9.4 and 9.9. This agreement is not too good, but perhaps is within the limits of error of the measurements. The observed increase in δ as λ is decreased from 3.18 cm to 1.24 cm if real, must be attributed to the effect of the absorption lines at higher frequencies. As the frequency is increased from zero upward, the δ value of NH_3 should first increase above the static value and then decrease as the microwave absorption band is approached. The observed values are both lower than the static value; the meaning of this result is uncertain.

ACKNOWLEDGMENTS

We wish in particular to express our indebted-

¹⁴ C. P. Smyth, *Dielectric Constant and Molecular Structure* (Chemical Catalog Company, New York, 1931).

¹⁵ C. P. Smyth and R. H. Wiswall, Jr., *J. Chem. Phys.* 9, 156 (1941); C. P. Smyth and K. B. McAuliffe, *J. Chem. Phys.* 1, 490 (1933); and *J. Chem. Phys.* 2, 499 (1934).

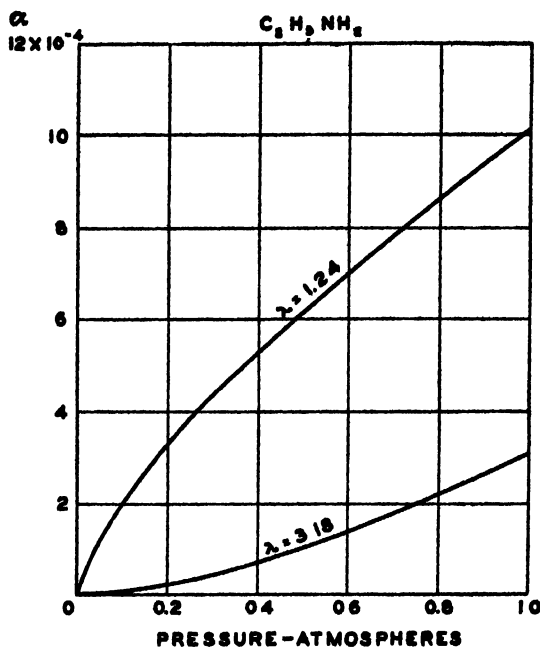


FIG. 6. Absorption of ethylamine.

ness to George W. Leck, Jr., for invaluable help in building microwave generating and power measuring equipment and to Elizabeth T. Bush for constructing wave guide components, preparing gases in the laboratory not readily available commercially, and taking many measurements. Also we express our appreciation to Irving Wolff of these Laboratories for support and encouragement, and to John Turkevich of Princeton University, for critical and helpful discussions during the course of the investigation.

Note added in proof, July 9, 1946.—Figure 4 in the paper referred to in reference 1 shows the absorption of ammonia at reduced pressure at three different wave-lengths. The information in this figure is quite consistent with the map of the fine structure of ammonia published by Bleaney and Penrose in the March 16, 1946, issue of *Nature*. The curve in Fig. 4 at 1.243 centimeters rises steeply at low pressure, but, after a few millimeters of pressure are reached, bends over and rises less steeply. A curve of this character indicates that the observations were taken near the maximum of one of the spectral lines. True enough, Bleaney and Penrose map an unresolved pair of lines at this wave-length to which they assign the quantum numbers (4,4) and (10,9). The curve at 1.320 centimeters also exhibits a point of inflection and these observations were taken on the complex and unresolved absorption line designated (5,4), (4,3), and (6,5). By contrast, the experimental curve at $\lambda = 1.227$ is linear which is consistent with the fact that this wave-length falls between the (4,4) and (5,5) lines.

Electrical Observations of the Austenite-Martensite Transformation in Steel

ANDREW W. McREYNOLDS
*Naval Proving Ground, Dahlgren, Virginia**
(Received March 14, 1946)

A method is described for observing the austenite-martensite transformation in steel by measuring electrical resistance. Small wire samples are heated and cooled either in vacuum or in air while temperature and resistance are automatically recorded. Temperature throughout the sample is uniform enough to permit a determination of resistance as a function of temperature. It is shown that from these data a complete record of transformation during cooling can be computed. Such transformation data have been obtained on three steels of widely different compositions, representing the types of steel in which transformation is most difficult to observe.

I. INTRODUCTION

AT temperatures above about 750°C, the stable phase of steel is the face-centered austenite. Relatively slow cooling causes a redistribution of carbon by diffusion and a readjustment of the lattice to the body-centered structure stable at lower temperatures. If cooling is rapid enough, however, the steel will instead transform without diffusion to a metastable tetragonal structure known as martensite, which is much harder than the products of transition by diffusion.

Study of the austenite-martensite transition is important in ferrous metallurgy, but its properties are such that observation is experimentally difficult. Some of the properties on which there is general agreement are:

1. Plate-like crystals of martensite form almost instantaneously, probably by a shearing process, with no diffusion.
2. The amount of transformation which has taken place at each temperature during cooling is a function of that temperature alone and is independent of the rate of cooling.
3. Transformation begins only when a definite temperature (referred to as the M_s point) has been reached, and this point cannot be changed by increasing the cooling rate.
4. The M_s point is depressed, approximately linearly with the amount added, by the addition of alloying elements.

Two general experimental methods can be applied to quantitative observation of the trans-

formation: metallographic examination after heat treatment, or direct observation during transformation by simultaneously measuring temperature and some physical property connected with the structure. The latter method requires that temperature be uniform over at least a portion of the sample, and that some physical property of that portion be measured continuously during cooling and correlated with the structure.

Bain and Davenport¹ first used the metallographic method in 1930 to study isothermal transitions to pearlite or bainite, and to derive their well-known "S" curve. Transformations to martensite could not be observed by that method, but in 1940 Greninger and Troiano² slightly modified the technique of heat treatment by interrupting quenching of the sample at some temperature T with a brief reheating to temper the martensite formed above temperature T , and render it distinguishable from that formed below T . Their results indicated the properties of the transition listed above. The modified S curve based on those properties is shown in Fig. 1. Payson and Savage³ applied the same "quench-temper" method to a study of the effect on the M_s point of alloying elements in steel.

Direct measurements have been made by a number of investigators using such phenomena as thermal arrest, dilatation, magnetic changes

¹ E. S. Davenport and E. C. Bain, *Trans. A.I.M.E.* 90, 117 (1930).

² A. B. Greninger and A. R. Troiano, *Trans. A. S. Metals* 29, 537-574 (1940).

³ P. Payson and C. H. Savage, *Trans. A. S. Metals* 33, 261-275 (1943).

* This article comprises part of a dissertation submitted to the University of California in partial fulfillment of the requirements for the degree of doctor of philosophy. Experimental work was carried out at the Naval Proving Ground, Dahlgren, Virginia.

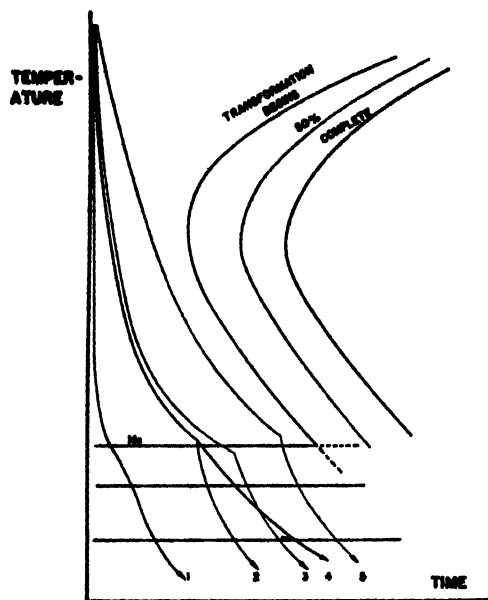


FIG. 1. Typical transformation S curves for steel and cooling curves for test specimens. (1) Rapid gas quench, (2-4) initial cooling in vacuum, gas cooling through transition region, (5) slow initial cooling.

and the noise of structural change. These measurements have been limited, however, to determination of the M_s point only, or to measurements at and below room temperature, on steel of such composition that it remained mostly austenite until cooled below room temperature.

Temperature-time measurements on small samples quenched under various conditions were made by French and Klopsch,⁴ Esser, Eilender, Riedel, and Spenle,⁵ Wever and Engel,⁶ Wever and Rose,⁷ and by Greninger.⁸ Temperature was measured by galvanometers with periods of 0.020 second or less. The M_s point as determined by the thermal arrest was found independent of the cooling rate in all cases.

Tamann and Scheil⁹ and later Scheil¹⁰ made dilatometric measurements in the temperature

⁴ H. J. French and O. Z. Klopsch, Trans. A. S. Steel Tr. 6, 251-293 (1924).

⁵ H. Esser, W. Eilender, K. Riedel, and E. Spenle, Archiv. f. Eisenh. 4, 113-144 (1930).

⁶ F. Wever and N. Engel, Archiv. f. Eisenh. 5, 377-382 (1932).

⁷ F. Wever and A. Rose, Mitt. Kaiser Wilh. Inst. f. Eisenf. 19, 289-298 (1937).

⁸ A. B. Greninger, Trans. A. S. Metals 30, 1-26 (1942).

⁹ G. Tammann and E. Scheil, Zeits. f. anorg. Chemie 133, 1-21 (1926).

¹⁰ E. Scheil, Zeits. f. anorg. Chemie 183, 98-120 (1929).

region below 20°C on high carbon steel. They found transformation to depend only on temperature. Förster and Scheil¹¹ made both acoustical and electrical resistance measurements in the same temperature region and found the time of formation of a single martensite plate to be of the order of 0.02 second.

All of the above methods of direct observation were limited to rather special conditions and were not adequate to determine complete data on the progress of transformation. The present paper describes a method for recording electrical resistance and temperature and for interpreting those data to furnish a complete record of the amount of transformation at all stages of the cooling. The technique of heating and cooling was such that at any instant the temperature throughout the sample was uniform within a few degrees. Electrical resistance is a suitable physical property to provide a measure of the proportions of the constituents present because it can be easily and rapidly measured and recorded at all temperatures, and because, as will be shown below, it can be quantitatively correlated with structure. The advantages of the electrical method over metallographic methods are:

1. It yields precisely measurable data and requires no visual estimates of relative proportions of constituents.
2. Measurements can be made directly during a normal cooling operation.
3. It is applicable to steels of almost any composition.
4. A complete record of percent of the sample transformed vs. temperature can be determined from the cooling of one specimen.

II. APPARATUS

Two different types of apparatus were used, one for heating and cooling the samples in vacuum and one for heating and cooling in fluid media. The first method has the advantages of more precise control of temperature and freedom from surface contamination or decarburization of the sample. The second method has the advantages of greater ease in preparing and mounting samples and a wider range of applicability to different compositions of steel. Since

¹¹ Förster and E. Scheil, Zeits. f. Metallkunde 28, 245-247 (1936).

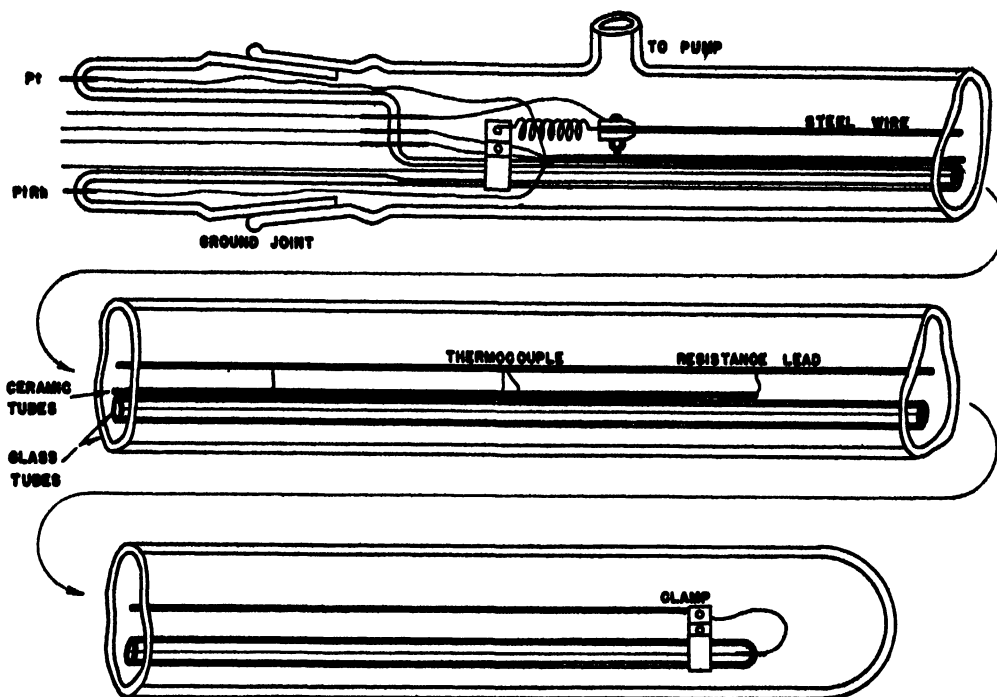


FIG. 2. Vacuum chamber, diameter 1.5 in., and steel wire sample.

cooling can be accomplished much more rapidly in fluid media than in vacuum, the second method is applicable to steels which require rapid cooling to cause transformation to martensite.

1. Apparatus for Heat Treatment in Vacuum

The samples were cut from heavy stock and drawn into uniform circular wire of 0.026-inch diameter. This size was chosen as small enough to allow cooling at rates up to 150°C per second without excessive radial temperature gradients and at the same time large enough to allow the attachment of 0.002-inch thermocouple leads without greatly altering the metal temperature near the junction. Segments of wire 50 cm long were polished with fine emery cloth, cleaned with carbon tetrachloride, and suspended under slight spring tension between clamps mounted on a 7-mm glass tube (see Fig. 2). The complete sample holder was inserted in a larger glass tube evacuated to 10^{-5} mm Hg. Heating was accomplished by passing 60-cycle alternating current directly through the sample. Cooling at rates of from 1°C per second to 200°C per second could

be accomplished either by radiation alone, or by transfer of heat through a small amount of helium, of the order of 1-mm Hg pressure, admitted for that purpose. The wire surface was carefully polished since the heat emission constant of the surface had to be uniform to preserve axial uniformity of temperature.

Temperature measurements were made by the use of a thermocouple of two 0.002-inch wires, platinum and platinum-10 percent rhodium, spot-welded to opposite sides of the sample at its center. (Fig. 3.) The thermocouple leads could be connected by a selector switch to either a Rubicon Precision Type B potentiometer or a galvanometer. The galvanometer was a Leeds & Northrup Type R, with 500-ohms coil resistance, which had been modified by the installation of a 0.002-inch round platinum-10 percent rhodium suspension to have a period of about 1 second. Its deflections, recorded photographically on a drum of sensitized paper moving at 1 cm per second, were found to be reproducible within 0.1 mm, corresponding to about 1°C. (A rapid recording potentiometer of the "Spec'lomax" type would be satisfactory for temperature recording.)

Resistance was measured between two 0.002 inch platinum-10 percent rhodium wires, one spot welded to the sample 7.5 cm on each side of the thermocouple leads. Measurements were thus taken only on the central 15 cm of the wire, and 17.5 cm were allowed on each end to eliminate non-uniformity of temperature due to conduction to the supports. Temperature uniformity was verified before each run by measuring the thermoelectric voltage between each of the resistance leads and the central platinum thermocouple wire and comparing with the voltage of the central thermocouple. To measure resistance, a constant direct current of 0.5 amp was passed through the wire and a galvanometer of the same type as used for temperature measurements was connected across the 0.002-inch resistance leads. Resistance and temperature measuring galvanometers recorded on the same moving film. A calibration was recorded on each run by connecting the resistance-measuring galvanometer across a standard resistance in series with the sample, and timing lines were marked on the paper by a lamp flashing at 1-second intervals.

Procedure in making a run was: The sample was heated in vacuum to 900°C and held at that temperature for 10 minutes to permit all carbon to go into solid solution. (A longer time was found to have no effect on the M_s temperature.) The heating current was then cut off to start the cooling. The time required in vacuum to reach the transition temperature (260°) was about 140 seconds. This time could be reduced

to as little as 4 seconds by the admission of a little helium to the system, or could be made as long as desired by reducing the heating current gradually. A small quantity of helium was always admitted to the system when the temperature was a few degrees above or a few degrees below the M_s point, in order to accelerate cooling through the transition region since cooling by radiation alone would have been extremely slow at low temperatures. The time required to cool through the transition temperature range was thus made too short to allow either the formation of diffusion products (e.g., bainite) from the austenite or an appreciable tempering of the newly-formed martensite. After the sample had cooled to room temperature, the chamber was again evacuated and the sample reheated to 200°–250°C in a time of 30 seconds. Since resistance could not be measured with the heating current on, heating was carried out in steps of about 20°, with the heating current turned on for one second and then off for three seconds while the measurements were made.

Both temperature and resistance were continuously recorded during the cooling and reheating operations. A typical record is shown in Fig. 5. Each sample was used for about ten records under various cooling conditions and was then divided into sections for carbon analysis, microscopic examination, and hardness tests.

2. Apparatus for Heat Treatment in Fluid Media

Bars 3 to 4 inches long and varying in cross-sectional size and shape from 0.025-inch diameter circles to 0.060-inch squares were cut from the sample material. Each bar was bent into a U-shape and supported by 36-gauge Chromel and Alumel wires spot welded near the ends of the U as shown in Fig. 4. The one Chromel and five Alumel wires served both as supports and as current, potential, and thermocouple leads. Each wire ran through a separate hole in a bundle of ceramic tubes and connected to one element of a flexible 6-conductor cable leading to the recording apparatus. The same recording instruments as described in the preceding section measured the thermoelectric voltage and the potential drop across a 1-inch segment of the sample caused by direct current of 0.5 ampere

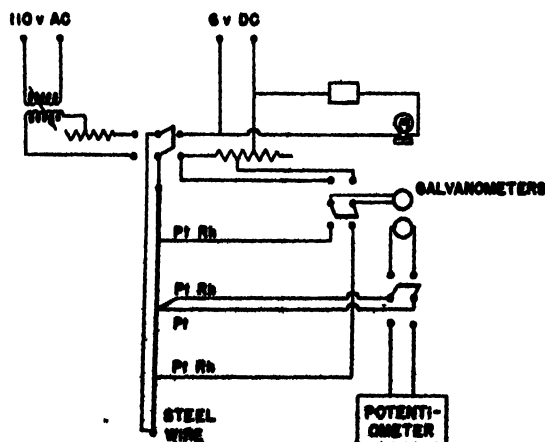


FIG. 3. Electrical circuit diagram.

which was passed through the sample. Thermocouple leads were attached outside the segment through which current passed, in order to avoid spurious potential drop due to the measuring current.

Heating was accomplished by inserting the sample (but not the ceramic tube support) either into a small lead bath 1 inch in diameter by 8 inches deep, heated to 900°C in an electric furnace, or into a hollow ceramic tube $\frac{3}{4}$ inch in inside diameter by 10 inches high, which was closed at the bottom and heated in the same furnace as the lead bath. After remaining at 900°C long enough for carbon to go into solid solution, the sample was cooled by withdrawing it into still air, or if rapid cooling was required, into a blast of air from a high speed fan mounted just above the heating furnace. After initial cooling by air blast, the sample was removed to still air to cool more slowly through the transition region. This cooling procedure was found to be rapid enough to quench a 0.025-inch round sample of plain carbon (0.88 percent) steel wire from 900°C without allowing any observable transformation at temperatures above the M_s point of 200°C. Although in that case time to cool from 900°C to 300°C was only 2.5 seconds, satisfactory temperature and resistance records were obtained during slower cooling below that temperature. Cooling rates at the beginning point of transformation were of the order of 20°C per second for the smallest samples used.

An alternate cooling method tried was to withdraw the sample from the heating furnace, immediately plunge it into a small bath of molten NaNO_2 and KNO_3 at 300°C, then allow it to cool with the salt bath while continuously stirring. This procedure was found to have no advantages over air cooling and to be much less convenient.

Some samples were cooled in air down to 120°C and then immersed in boiling water (at 100°C to restore complete uniformity of temperature. They were next immersed alternately in water at room temperature (27°C) and at 100°C several times, being left in each bath long enough for a resistance measurement to be taken. Data thus obtained provided an accurate measurement of the amount of transformation between 100°C

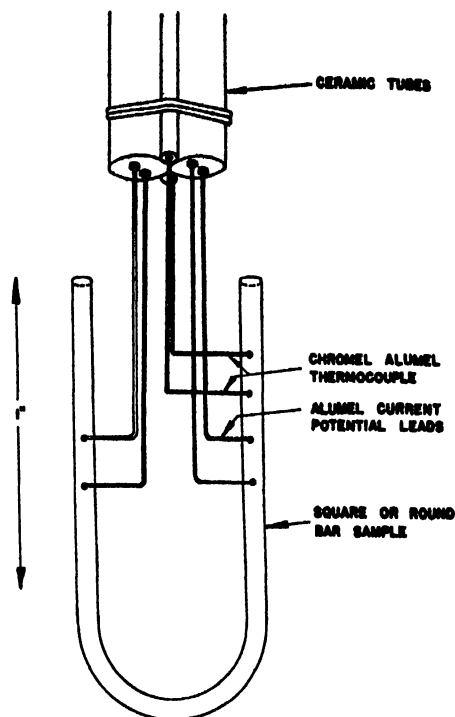


FIG. 4. Sample mounting for heating in furnaces and liquid baths and cooling in air.

and 27°C, and of the temperature-resistance relation of the final structure. The method of interpretation of data is explained in the following section.

III. INTERPRETATION OF DATA

A sample photographic record, obtained on cooling in vacuum and helium, is shown in Fig. 5. Galvanometer deflections were read to an accuracy of 0.1 mm, corresponding to 0.1–0.3 percent of the resistance and to about 1°C in temperature. Thermocouple deflections were converted to degrees C, but resistance measurements were left in arbitrary units. The data thus obtained could be interpreted to determine the amount of transformation at any stage of the cooling and also the heat of transformation.

The data were plotted (see Figs. 6–8) as resistance-temperature curves which exhibit a sharp break in slope at the M_s point. An essential condition of the procedure was that temperature must be uniform throughout the sample during cooling, since even a slight non-uniformity could give a spurious shape to the resistance-tempera-

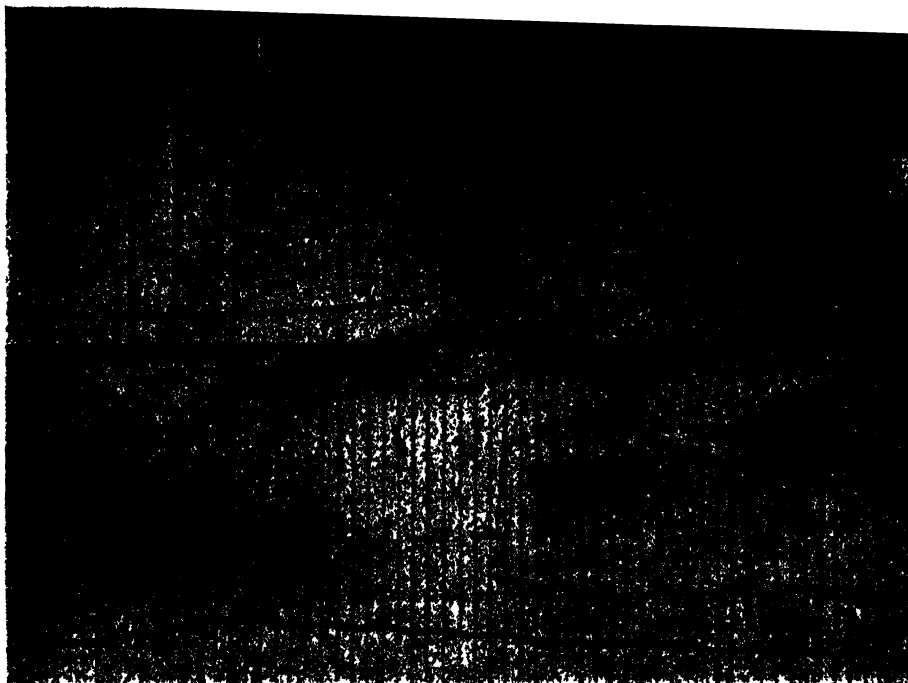


FIG. 5. Typical record of temperature and resistance. Top portion represents cooling and bottom portion reheating.

ture curve. The sharpness of the break at the M_s point could, however, be used as a criterion of uniformity, and all records in which the break was blurred into a round corner were discarded.

The curves in Figs. 6-8 will be seen to consist of two linear portions and a connecting curve. The two linear segments represent the resistance-temperature relation when the crystal structure is 100 percent austenite and 100 percent martensite, respectively, and the connecting curve represents the resistance as a function of temperature of a mixture of the two constituents as their proportions continuously change with decreasing temperature. If it is assumed that the linear austenite resistance curve can be extrapolated below the M_s point without much error to show the resistance of austenite retained in that temperature region, then from the position of the connecting curve between the linear segments, the proportions of the constituents present at any time and temperature can be determined. The resistance of the mixture depends on both the amount and the distribution of the martensite plates in the austenite matrix. The extreme possible distributions are with the

martensite either entirely in series with the austenite or entirely in parallel. These distributions lead, respectively, to the relations:

$$\frac{\% \text{ Martensite}}{100} = (R_a - R) / (R_a - R_m) \text{ (series),}$$

$$\frac{\% \text{ Martensite}}{100} = \frac{R_a - R}{R_a - R_m} \times \frac{R_m}{R} \text{ (parallel),}$$

where

R = resistance of the mixture.

R_a = resistance of 100 percent austenite.

R_m = resistance of 100 percent martensite.

For a ratio $R_a/R_m = 1.5$, these relations differ by a maximum of 10 percent as shown in Fig. 9. Actually, the martensite is known to be distributed as randomly oriented small plates throughout the matrix. It was considered a sufficiently close approximation to assume that the martensite forms in thin parallel plates at 45 degrees to the direction of current flow and extending completely across the sample. It can easily be shown that with such a distribution $R = \frac{1}{2}(R_{\text{series}} + R_{\text{parallel}})$. On the basis of this

assumption, values of the function $(R_a - R)/(R_a - R_m)$ were computed and are shown in Fig. 9. The amount of transformation was determined from the resistance-temperature curves by first finding the ratio $X = R_a/R_m$, then using the appropriate curve from Fig. 9 to find the percent martensite corresponding to the value of $(R_a - R)/(R_a - R_m)$ at any desired temperature. A number of resistance-temperature-time curves were obtained on each sample, and for each curve Fig. 9 was applied to determine the temperatures at which transformation was 10, 25, 50, 75, and 90 percent complete.

Since the M_s point was independent of the rate of cooling and since microscopic examination showed the final structure to be martensite it is clear that the deviation of the curve from linearity represented a transformation, but it still had to be determined whether the shape of the curve was influenced by other reactions. The fact that no such reactions occurred above the M_s point is evidenced by the linearity of the resistance-temperature relation in that region. To investigate the possibility of tempering of

the newly-formed martensite below the M_s point, a number of samples were cooled in

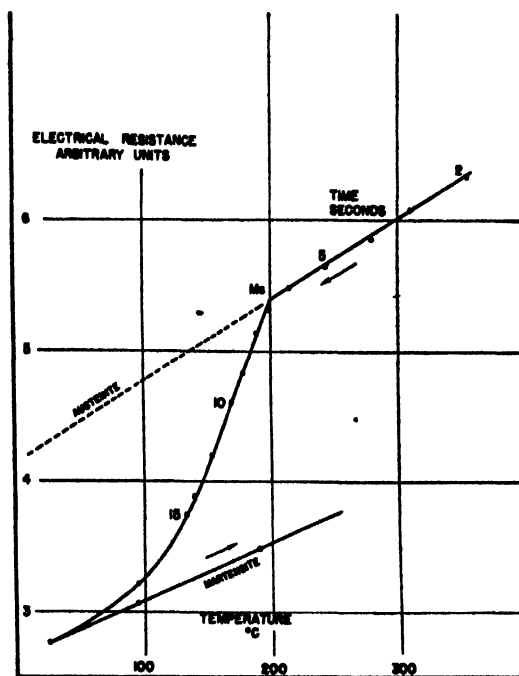


FIG. 7. Curve for 4.5 percent Mn low-carbon steel cooled rapidly in air blast.

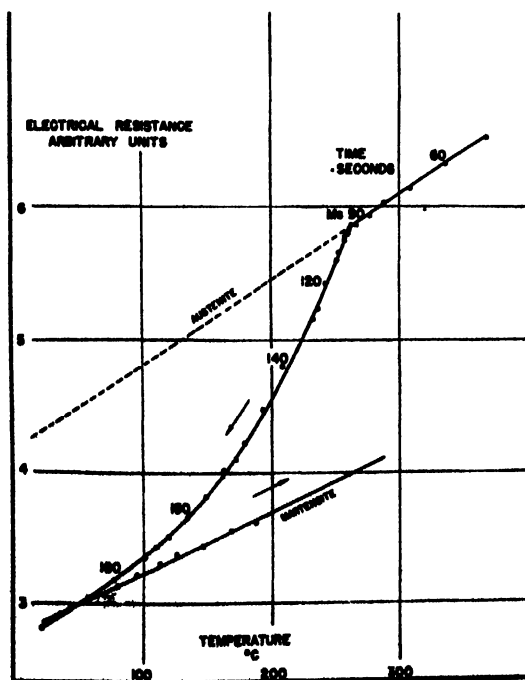


FIG. 6. Resistance-temperature curve for Ni-Cr alloy steel cooled in vacuum for 130 seconds, then more rapidly in helium. Position of non-linear segment indicates degree of transformation from austenite to martensite.

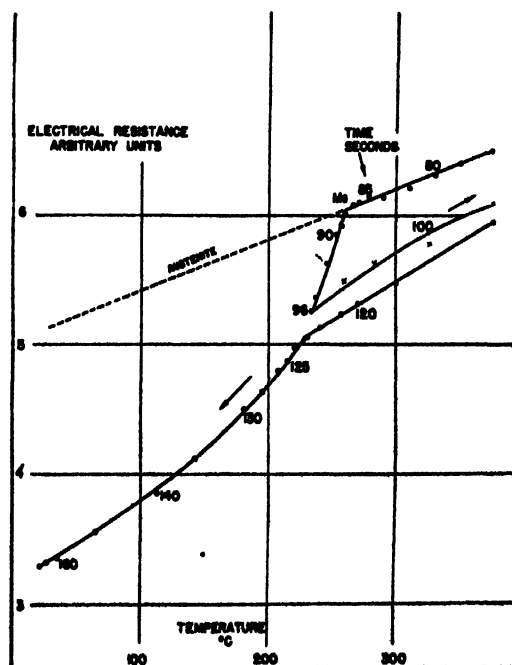


FIG. 8. Curve for Ni-Cr alloy steel. Cooling was interrupted at 130°C. Transformation stopped at that point and resumed when temperature next dropped below 130°.

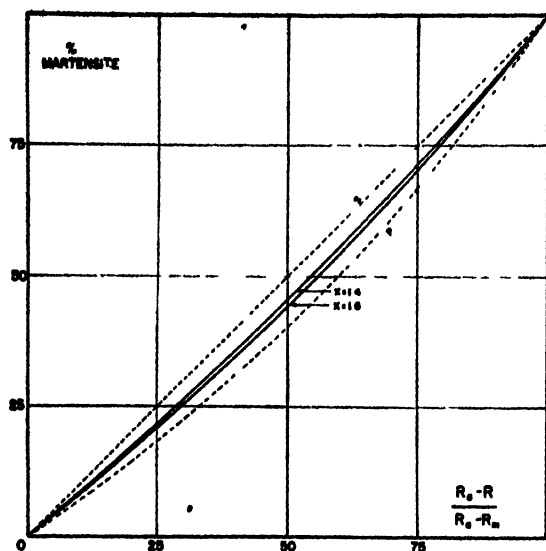


FIG. 9. Variation of resistance of an austenite-martensite mixture with proportions of the constituents. Dotted curves show variation with the constituents in series or in parallel with each other, solid curves with constituents assumed to be in alternate thin plates at 45° to the direction of current, $X = R_a/R_m$, R = resistance of mixture, R_a = resistance of 100 percent austenite, R_m = resistance of 100 percent martensite.

vacuum nearly to the M_s point, then at various rates through the transition region. Figure 10 shows that the computed amount of transformation as a function of temperature is the same for all of these tests. It is therefore evident that time-dependent tempering reactions either did not occur or did not influence the determination of the amount of transformation. Local stresses set up by the formation of martensite plates might be expected to increase the resistivity of both the martensite and the surrounding austenite and thus influence the shape of the resistance-temperature curve. The amount of the increase in resistivity, however, should be not greater than would be caused by extensive cold work; that is around one percent for steel. Furthermore the increases in resistivity of austenite and martensite enter into the determination of the amount of transformation in such a way as to approximately cancel each other. It is therefore safe to assume that stresses accompanying transformation introduce no serious error.

Metallographic examination of the samples after cooling to 20°C showed them to be all martensite, as nearly as could be determined. The resistance curves also indicate that transfor-

mation had practically ceased at that temperature. It has therefore been assumed in the computations that the structure after cooling to room temperature was 100 percent martensite.

From the data of Table II and from a determination of the decrease of rate of cooling as transformation begins, a value of the heat of formation of martensite can be calculated. The rate of heat radiation is substantially the same immediately before transformation begins as immediately after, but after reaction begins part of this heat is supplied by the exothermic reaction so that the cooling rate is reduced to a value θ_2 determined by the equation:

$$ms\theta_1 = ms\theta_2 + mHM\theta_2,$$

$$HM\theta_2 = s(\theta_1 - \theta_2),$$

(1)

$$H = \frac{s}{M} \left[\frac{\theta_1}{\theta_2} - 1 \right],$$

where

H = heat of reaction,

s = heat capacity of austenite,

m = mass under consideration,

M = fraction of martensite formed per degree decrease in temperature,

θ_1 = cooling rate just before transformation begins,

θ_2 = cooling rate just after transformation begins.

To determine H , only the ratio of cooling rates need be determined and therefore only the slopes of the temperature-time galvanometer record need be measured, and no calibration is required.

IV. RESULTS

Samples of the compositions listed in Table I were tested. Data on the progress of transformation with decreasing temperature are given in Table II and summarized in Table III. These

TABLE I

Sample Shape	0 0.026" wire	1 0.026" wire	2 0.060" wire	3 0.060" bar	4 0.025" wire
Element	Percent				
Carbon	0.41	0.41	0.38	0.09	0.88
Manganese	0.26	0.26	0.26	4.53	0.33
Molybdenum	—	—	—	0.93	—
Nickel	3.85	3.85	3.85	—	—
Chromium	1.85	1.85	1.85	—	—
Silicon	0.074	0.074	0.074	0.12	0.12
Phosphorus	0.015	0.015	0.015	0.014	—

results confirm a number of properties of the transformation which have been previously suggested on the basis of metallographic and other investigations.

The M_s point and also the amount of transformation which had taken place at a given temperature were found to be independent of both the over-all cooling rate and the rate of cooling through the transition region within the range investigated. Figures 10 and 11 show this property to hold for rates from 1°C per second to 100°C per second above the M_s point and for times between 10 and 75 seconds to cool from the M_s point to the point of 90 percent transformation.

Tests in which cooling was interrupted before the completion of transition showed that transformation stopped as soon as cooling was interrupted but resumed when cooling next proceeded below the point of interruption. Figure 8 is the record of one such test in which the sample was cooled to about 230°C , rapidly reheated to about 360°C , then allowed to cool. It will be seen that the resistance curve does not return to its original position at 230°C , but the discrepancy is attributable to tempering of the martensite during reheating.

From the chemical compositions of the three samples values of the M_s points were calculated according to the approximate formulas of Payson and Savage,³ which were based on extrapolated metallographic data. The agreement between these predicted values and the experimental data, both shown in Table III, confirm that the

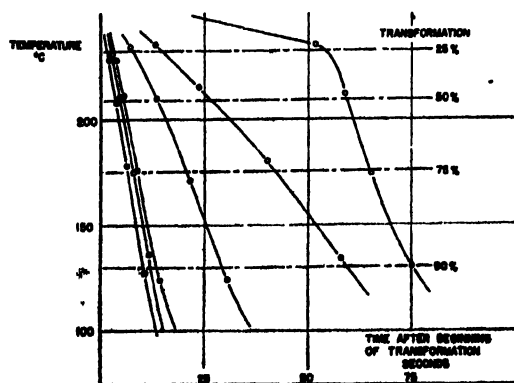


FIG. 10. Effect of cooling rate through the transformation region on the progress of transformation. Cooling curves are for Ni-Cr alloy steels.

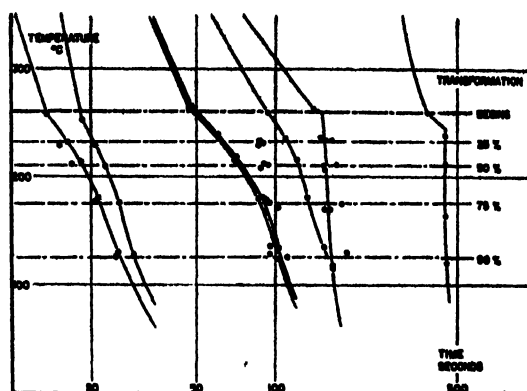


FIG. 11. Effect of cooling rate on the progress of transformation. Cooling curves are for a number of Ni-Cr alloy steel samples. Points represent individual determinations from the resistance-temperature curves of the temperature at which transformation reached various stages of completion.

M_s point is determined almost entirely by the chemical composition.

Although no tests were conducted for the specific purpose of measuring heat of formation of martensite from austenite, values of this quantity were determined from some of the records by using Eq. (1). Values of $s=0.115$ cal. per gram per degree centigrade and $M=0.0108$ and $M=0.0089$ per degree centigrade for samples 0 and 1, respectively, were used in the calculations.

V. CONCLUSIONS

1. From measurements of electrical resistance and temperature, continuously recorded by the method described, complete quantitative data on the transformation of a steel sample from austenite to martensite can be determined. The electrical resistance method is applicable to almost any type of steel, including the two types experimentally most difficult to study:

- Steel of low carbon content (less than 0.4 percent) which cannot be studied metallographically or by any other method yet developed.
- Rapid-reacting steels which must be quenched very rapidly to produce the martensite structure.

Satisfactory results have been obtained on both a 0.09 percent carbon steel and a 0.88 percent carbon plain carbon steel.

TABLE II. Transformation temperatures.

Sample	Record	Initial temp. to M_s point	Cooling medium M_s point to room temp.	Begins (M_s)	10%	Transformation 25%	50% °C	75%	90%
0	1	vacuum	helium	257	247	234	210	171	124
0	2	vacuum	helium, rapid	—	—	228	208	178	127
0	3	vacuum	helium, slow	258	248	235	215	180	134
0	4	vacuum	helium, rapid	—	—	231	211	175	136
0	5	helium, rapid	helium, rapid	—	—	228	211	176	124
0	6	vacuum	helium, slow	261	249	235	212	174	130
0	7	vacuum, slow	helium	258	250	238	212	163	118
1	1	vacuum	helium	262	247	229	194	147	100
1	2	vacuum	helium	264	249	227	200	150	95
1	3	helium, slow	helium	263	249	232	203	158	102
1	4	helium, slow	helium	263	245	229	200	152	95
1	5	vacuum	helium	264	253	237	209	168	114
1	6	vacuum	helium	264	251	234	207	168	114
1	7	helium, rapid	helium, rapid	250	241	229	208	173	126
1	8	helium, rapid	helium, rapid	256	246	232	211	178	128
1	9	helium, slow	helium	264	252	239	217	179	133
1	10	helium, slow	helium	262	249	234	215	178	123
Average				260	248	232	209	169	119
2	1	lead-salt, rapid	salt-slow	277	267	255	232	200	165
3	1	air, rapid	air, rapid	342	333	317	281	229	174
3	2	air	air	336	329	316	280	216	154
3	3	air, slow	air, slow	328	319	304	272	207	138
3 Average				335	327	312	278	217	155
4	1	air blast	air	200	189	173	150	123	93
4	2	air blast	air	192	181	167	147	121	99
4 Average				196	185	170	148	122	96

TABLE III.

Sample	0, 1	3	4
Composition	0.41C 3.85Ni 1.85Cr	0.09C 4.53Mn	0.88C 0.33Mn
Transformation: Begins	260°C	335°C	196°C
25%	232	312	185
50%	209	278	170
75%	169	217	148
90%	119	155	96
M_s point calculated from composition	242°C	315°C	210°C

TABLE IV.

Sample	Cooling rate above M_s °C per second	Cooling rate below M_s	Heat of transformation cal. per g
0	21.0	7.1	25.2
0	12.8	5.25	18.6
0	21.5	9.3	16.9
0	22.1	8.7	20.8
1	26.2	11.05	17.6
1	29.3	11.5	20.0
Average			19.8
3	9.7	3.17	21.8
3	13.5	3.26	33.1
3	14.9	5.5	21.1
3	11.05	3.66	21.4
3	10.85	3.42	22.8
3	17.7	4.32	32.8
3	15.8	4.15	29.5
Average			24.5

2. Data on the alloy steel samples show that the temperature at which transformation begins and the amount which has been transformed at any temperature during cooling are independent of the time required for the steel to cool to that temperature (if no diffusion reactions occur), and hence, if cooling is interrupted before transformation is complete, the formation of martensite ceases and resumes only when the temperature is reduced below the point of interruption. These properties of transformation are at present

generally accepted on the basis of previous direct measurements (during slow cooling below room temperature) and of metallographic observations.

The author is indebted to Mr. H. Elmendorf of the American Steel and Wire Company for drawing the samples into wire form, and to

Professor L. B. Loeb of the University of California, and Commanders B. R. Queneau, R. A. Sawyer, and R. N. Varney, of the Naval Proving Ground for helpful discussions. The opinions

and assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

Energy Build-Up in Magnetrons

LLOYD P. HUNTER

Westinghouse Electric Corporation, Pittsburgh, Pennsylvania

(Received April 3, 1946)

An analysis of the operation of resonant cavity magnetrons is made without regard to the mechanism of conversion of the d.c. input power into the r-f output power. The magnetron is represented by its equivalent circuit and is assumed to be essentially in equilibrium at every point of its energy build-up curve. The last assumption is shown to be reasonable on the basis of measured starting times. From the data of load impedance charts and the rise of the r-f pulse, the relation between power generated and vane voltage is deduced. By considering the division of the energy between that stored in the resonant system, and that dissipated in the load, the law of build-up is derived. The dependence of starting time on load is calculated and agrees with experiment when the extra energy stored in the connecting line is taken into account. The starting time is affected slightly by the initial noise level and becomes infinite below a minimum Q . For high Q values the starting time can be varied only by changing the energy stored in the line, which may enable one to design a line for "mode" suppression.

INTRODUCTION

ONE of the most important parameters of magnetron performance is the radiofrequency voltage on the slots of the magnetron. This parameter is not directly accessible to measurement and has therefore been little considered in magnetron design. The purpose of this paper is to deduce an empirical law relating the generated power of a magnetron to the slot voltage, and then to show how this law may be used to calculate various features of magnetron performance which are of practical interest.

This investigation is specifically applied to resonant cavity magnetrons but illustrates a method of performance analysis which should be quite valuable when applied to any of a large class of high frequency generators. The purpose of the method is not to give any theory of the conversion mechanism in which the d.c. applied energy is converted into radiofrequency energy, but rather to give a performance analysis which should contribute greatly to the understanding necessary for intelligent engineering design of oscillators of the magnetron type.

First the slot voltage will be calculated for

various values of load with a constant value of d.c. voltage applied to the magnetron. This will enable us to plot the output power delivered for these various loads against the slot voltage. Next we will supplement this curve of power output against slot voltage with similar data obtained from curves of the build-up of the r-f energy at the beginning of an applied d.c. voltage pulse. The curve will now extend over a wide range of slot voltages and we will proceed to fit an empirical relation to it.

The empirical relation so deduced will be used first to integrate the differential equation of energy build-up. The resulting equation can then be solved explicitly for the time, and we will proceed to calculate the variation of starting time as a function of load and of noise level. The form of the r-f build-up curve will also be calculated along with the conversion efficiency of the magnetron.

SECTION I

Relation of Power Generated to Slot Voltage

In this section the one fundamental assumption involved is that steady-state data, such as

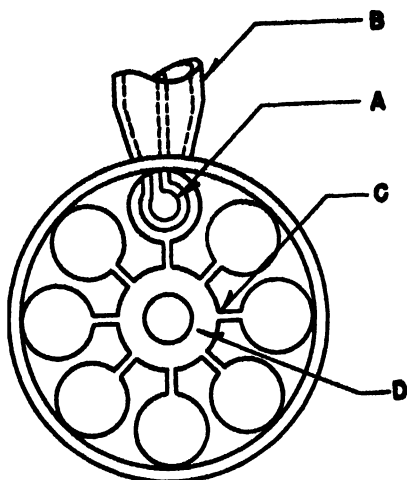


FIG. 1. Resonant cavity magnetron.

the variation of output power with load, may be applied to the build-up process. The results of starting time¹ measurements support such an assumption since they show that the complete build-up process occurs in about 300 cycles of the r-f. This means that the voltage is essentially constant for a few cycles and we are justified in postulating a state of quasi-equilibrium for our calculations.

In Fig. 1 is shown a diagram of a resonant cavity magnetron of the so-called "hole and slot" type. At A is shown the coupling loop which bleeds off r-f power and sends it down the coaxial line B to the load. At C is shown one of the slots across which the r-f voltage is developed by the circulating electron cloud in the interaction space D. Such a magnetron is typical, and although it may have a wave guide output rather than a coaxial line-coupling loop arrangement, it will serve for our discussion in either case.

One well-known method² of representing the performance of a magnetron is to plot contours of constant power out in the complex load impedance plane while holding the d.c. anode voltage constant. Such a load impedance chart is shown in Fig. 2. The solid contours are constant power contours while the dashed contours are constant operating frequency contours. From such data it is possible to calculate the manner

in which the power output and power generated vary with the slot voltage. However, only the variation at high slot voltages is available in these data and one must look elsewhere for this variation near zero slot voltage. It is the purpose of this paper to show that such information can be obtained from measurements¹ of the build-up of the r-f envelope at the beginning of a pulse. To do this we will first examine the available data concerning the variation of output power with load, and then we will show how the measurements of r-f build-up allow us to extend these data.

First calculating the variation of output power with r-f voltage on the slots of the magnetron, we will assume a simple series equivalent circuit for the magnetron as seen from the back of the magnetron cavity. Now by definition

Transmitted $Q = Q_T$

$$= \frac{2\pi \text{ energy stored}}{\text{energy out/cycle}} = \frac{2\pi V^2/\omega z_0}{P_0/f}, \quad (1)$$

giving

$$V = (P_0 Q_T z_0)^{1/2} \quad (\text{r.m.s. volts}),$$

where P_0 = output power and $z_0 = (L/C)^{1/2}$ for the magnetron equivalent circuit.

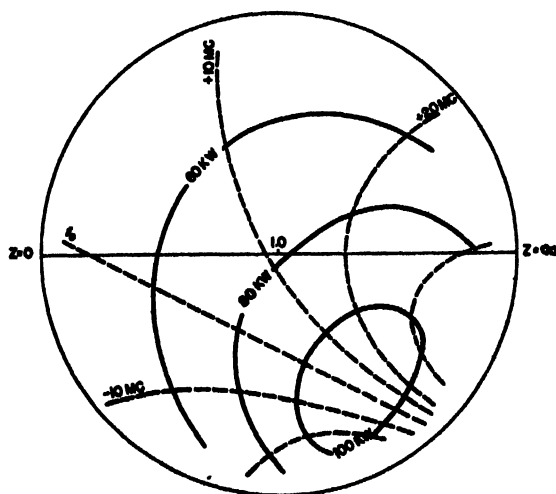


FIG. 2. Load impedance chart.

This formula gives us the relation between the r-f voltage on the slots of the magnetron and the power delivered to the load. We will not concern ourselves at this point with the power lost in the circuit, since such losses do not enter into the

¹ More and Miley, Massachusetts Institute of Technology Report 509.

² Evans and Rieke, Massachusetts Institute of Technology Report 52-2.

calculation of the slot voltage using Eq. (1), nor will they enter into the calculation of the build-up of stored energy with time, since that may be approached by using the output power as well as the power generated. We will consider all losses, however, when we come to calculate the total power generated as a function of slot voltage.

Referring the load impedance chart of a 2J32 tube³ to constant d.c. voltage by means of a performance chart, we may calculate the curve of Fig. 3 by means of formula (1). Values of P_0 and corresponding Q_r values may be obtained directly from the load impedance chart if it is remembered that the Q varies directly with the voltage standing wave ratio, and that the region of low Q or heavy loading corresponds to the region of crowded frequency contours. Figure 3 is a plot of output power as a function of the square of the r-f voltage on the slots of the magnetron. As can be seen from Fig. 3, these data are not sufficient to tell us much about the low voltage end of this characteristic curve. However, the results¹ of More and Miley on starting time and build-up of the r-f pulse in magnetrons is sufficient to indicate how the curve of Fig. 3 approaches zero.

In Fig. 4 is shown a typical family of r-f

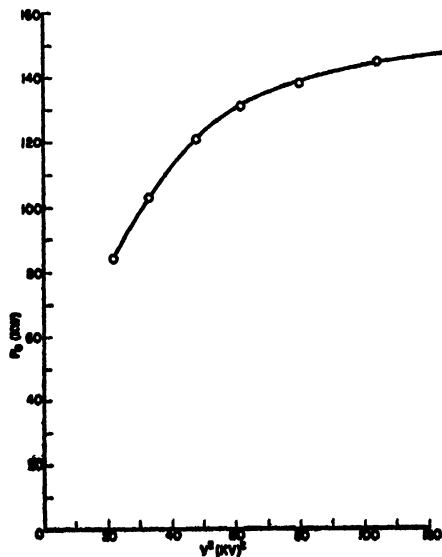


FIG. 3. Characteristic curve from load impedance chart.

³ Platzman, Evans, and Rieke, Massachusetts Institute of Technology Report 451.

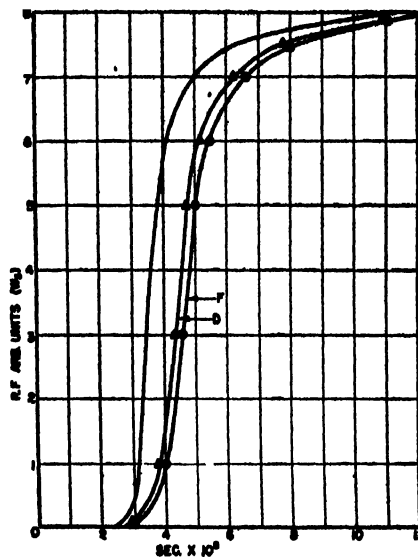


FIG. 4. Build-up curves from r-f envelope.

build-up curves. The curve marked d is plotted directly from the photograph d of Fig. 3 of reference 1. Similarly, curve f is plotted from photograph f . These two curves are for different rates of rise of the voltage pulse given in reference 1. The solid curve of Fig. 4 represents the response of the magnetron to a perfectly square voltage pulse of zero time of rise. This curve was obtained from the other two by a method devised by Dr. William Altar,⁴ which gives the response of a general circuit to an ideal pulse of infinite rate of voltage rise, from the known response of the network to a pulse with a finite and constant rate of voltage rise. Since this method is an application of the inversion of Carson's theorem on a transient response, it presupposes proportionality of the r-f voltage response of a magnetron *vs.* applied d.c. voltage. Such a condition is more or less fulfilled in a magnetron once the threshold voltage V_0 necessary to maintain oscillations is reached, provided one considers the applied d.c. voltage to be $(V - V_0)$. In this application we are justified to some extent by the fact that the two experimental curves, so corrected, coincided exactly. The actual correction was made to the square root of the curves d and f , which, being curves of r-f power, are proportional to the stored energy in the circuit

⁴ William Altar, private communication.

rather than the voltage. The corrected curve was then squared to give the solid curve of Fig. 4.

Since the r-f power out is at every instant proportional to the stored energy, the curves of Fig. 4 may be thought of as curves of stored energy (W_s) against time. We may use the curve of zero time of rise of the voltage pulse to determine the rate of change of stored energy in the build-up process. The rate of change of stored energy must be equal to the difference between the total power generated and the power lost both in the circuit and in the load. Since at present we are concerned with extending the data of Fig. 3, by the use of the curve of Fig. 4, we need not consider the circuit losses, for although they are directly proportional to the square of the slot voltage, as are the stored

follows:

$$\frac{dW_s}{dt}(V^2) = f(V^2) - \frac{\omega W_s(V^2)}{Q_T}, \quad (2)$$

where W_s is stored energy, and $f(V^2)$ represents the net power generated as a function of the square of the slot voltage. Since the slot voltage completely determines the conversion process of the electrons, it will appropriately serve as independent variable. From the solid curve of Fig. 4 we obtain $dW_s/dt(V^2)$. To do this we measure the slope at various values of W_s and plot a curve of dW_s/dt against W_s , or what is the same thing, against V^2 . So far the W_s scale of the curve of Fig. 4 has been undetermined; however, it will now be determined from the equilibrium point which is the intersection of the load line with the curve of net power generated. The load line is the plot of the last term of Eq. (2) which for a constant frequency and Q_T is obviously a straight line through the origin in a P_0 vs. V^2 plot (see Fig. 5). Now we have a portion of the curve of net power generated as shown in Fig. 3, and, dashed, in Fig. 5; for once the stored energy is built up, the net power generated is equal to the power out. The intersection point of these two curves gives the operating point of the magnetron in the steady state, and thus fixes the total amount of stored energy in the magnetron. Having chosen a Q_T and a frequency, we may calculate W_s from the V^2 of the intersection point. Using this final value of W_s we can now determine the scale of W_s in Fig. 4 so that this value is reached as a final W_s value of all the curves of Fig. 4.

Having thus determined the absolute (or actual) values of W_s in Fig. 4 and hence the values of dW_s/dt we may now extend the curve of Fig. 3 by the use of our $dW_s/dt(V^2)$ obtained from Fig. 4. First drawing our load line

$$P_0 = \frac{\omega W_s}{Q_T} = \frac{V^2}{Q_T \tau_0}$$

(Fig. 5) and adding to it our curve of $dW_s/dt(V^2)$ we obtain the points on the curve of Fig. 5, which is our desired net power generated. That is clear from Eq. (2)

$$f(V^2) = \frac{\omega W_s(V^2)}{Q_T} + \frac{dW_s}{dt}(V^2). \quad (2a)$$

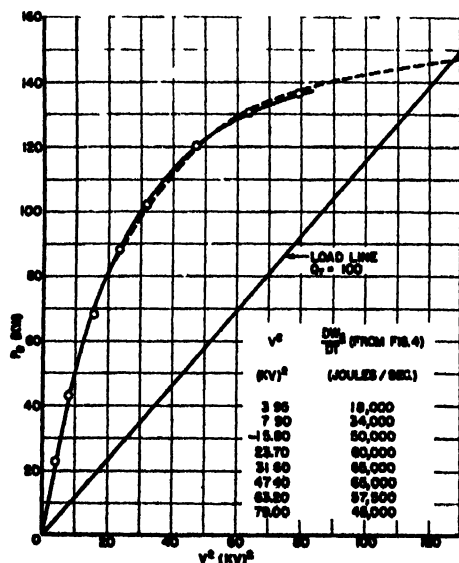


FIG. 5. Characteristic curve calculated from build-up curve.

energy, and the power out, each of these is independent of the others, and each may be considered an independent function of the square of the slot voltage. Thus we may consider the rate of change of stored energy equal to the difference between the net power generated and the load loss, where the net power generated is the total power generated minus the circuit losses.

The above relation may be expressed as

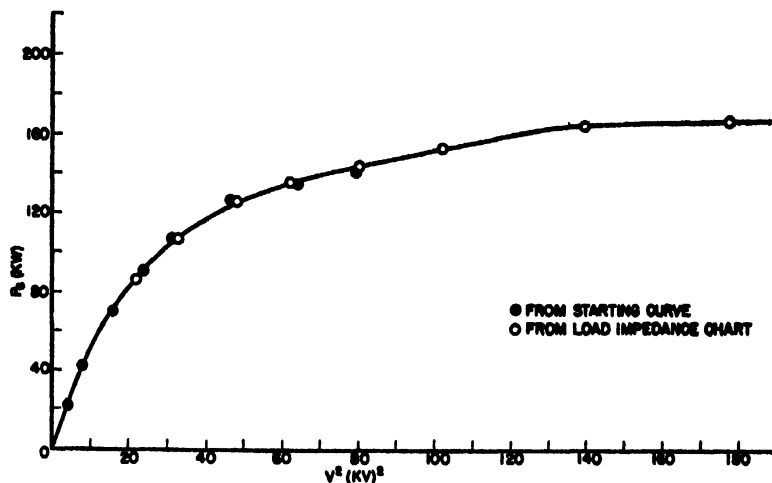


FIG. 6. Characteristic curve of 2J32 magnetron.

As stated above this curve is also the curve of P_0 vs. V^2 for steady-state operation. The curve of Fig. 3, obtained from the Rieke diagram is shown dashed in Fig. 5, and the solid curve is the curve obtained from the starting curve calculations. From the curve of Fig. 5 we may work backward and calculate starting curves for different load lines using the same graphical method described above. However, we shall first calculate the true curve of total power generated vs. slot voltage squared, and attempt to find an analytical expression for it before proceeding with any starting time calculations.

In order to determine the curve of total power generated vs. the square of the slot voltage, we must consider the circuit losses which we have so far found convenient and permissible to disregard. We will assume a constant unloaded Q_u for our magnetron as a preliminary step in the consideration of the circuit loss. The circuit efficiency now becomes

$$e = Q_u / (Q_u + Q_r)$$

and we may calculate the total power generated (P_g) from the output power (P_0) by using this circuit efficiency ($P_g = eP_0$). To do this numerically we must choose a numerical value for Q_u . Most S band magnetrons have unloaded Q 's lying between 1100 and 1800. Using a value of 1500 we may correct the curve of P_0 vs. V^2 of Fig. 5 to get the curve of P_g vs. V^2 given in Fig. 6. From this curve it can be seen that the power generated approaches zero with a finite slope.

This means that there is a limiting load beyond which the magnetron will not oscillate. This load corresponds to the load line which is tangent to the P_g curve at the origin, for this is obviously the last load line which can be drawn to intersect the curve of power generated.

Such a result is to be expected since it is clear that if the drain on the energy stored in the magnetron is so great that it effectively prevents the build-up of this energy above noise level, the magnetron will never be able to get started, and therefore will not oscillate. This is analogous to the conventional d.c. shunt generator in which the terminal voltage supplies both the exciting field and the load. There the power used by the load may so lower the terminal voltage that the exciting field will no longer be large enough to allow the generation of the power required by the load, and the generator will not operate.

Rather than continue with our graphical procedure we will attempt to fit an analytical expression to the curve of Fig. 6. This curve most obviously comes tangent to two lines, the limiting load line and the saturation power line. The limiting load line is a line through the origin with a finite slope, and the saturation power line is a straight line parallel to the V^2 axis. The simplest expression to fit such a curve is

$$f(V^2) = BV^2 / (1 + aV^2) = P_g, \quad (3)$$

which, for $V=0$, has the finite slope B , and for large values of V^2 approaches asymptotically $P_g = B/a$. In order to test the fit of this function,

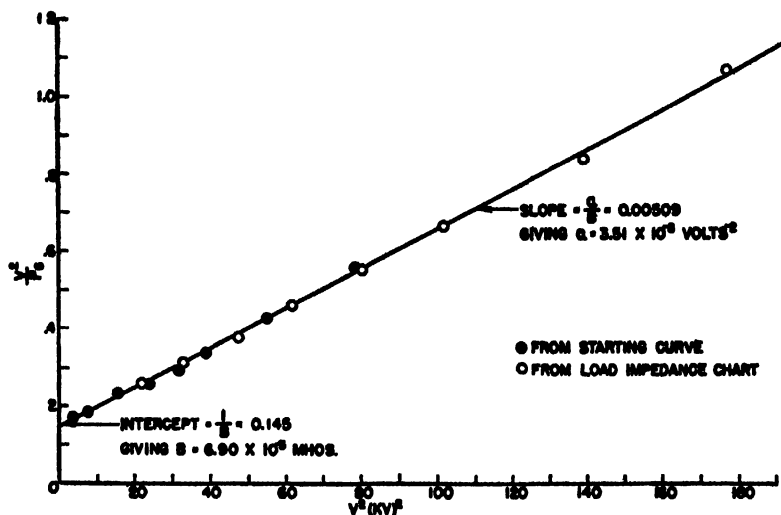


FIG. 7. Law of power generated.

we will write (3) in the form

$$V^2/P_0 = 1/B + aV^2/B. \quad (4)$$

This expression is a straight line in the coordinates V^2/P_0 and V^2 with an intercept $1/B$ and a slope of a/B . In Fig. 7 we see a plot of the data of Fig. 6 in the coordinates V^2/P_0 vs. V^2 . It is quite apparent that this plot is fitted by a straight line, which shows that the law proposed above does indeed represent the data of Fig. 6. This result is all the more remarkable when one considers that the data represented are from two different sources, the load impedance chart, and the starting curve. In Fig. 7 the data derived from the starting curve are represented by the closed circles while those derived from the load impedance chart are shown in the open circles.

SECTION II

Starting Time as a Function of Load

We now proceed to the integration of Eq. (2) using the $f(V^2)$ as found in Eq. (3). The result of this integration will then be solved explicitly for t and used to calculate the variation of starting time as a function of load.

If we now wish to include the circuit losses in Eq. (2), we find that it is best to combine them with the load and write, using Q_L instead of Q_T

$$\frac{dW_s}{dt}(V^2) = \frac{BV^2}{1+aV^2} - \frac{\omega W_s(V^2)}{Q_L}. \quad (5)$$

Assuming as is customary that W_s is the energy stored in the magnetron alone, we may write

$$W_s = AV^2,$$

where A is a constant equal to $1/\omega^2 L$ on our equivalent circuit basis. Equation (5) becomes

$$\frac{dW_s}{dt} = \frac{BW_s}{A+aW_s} - \frac{\omega W_s}{Q_L}. \quad (6)$$

Integrating we get

$$\frac{AQ_L}{Q_L B - \omega A} \ln(a\omega W_s) - \frac{Q_L}{\omega} \frac{Q_L B}{Q_L B - \omega A}$$

$$\times \ln(Q_L B - \omega A - a\omega W_s) = t + \text{const.}$$

Now in order to evaluate the constant we must assume a value for W_s at $t=0$. This corresponds to the initial noise level from which the oscillations build up. This value of W_s shall be represented by (db) decibels below the final value of W_s . Now we write the final value of W_s (W_s final) as

$$W_s \text{ final} = (Q_L B - \omega A)/a\omega$$

for at this value of W_s , t goes to infinity. This leads to a constant of integration

$$\text{const.} = \frac{AQ_L}{Q_L B - \omega A} \{ \ln(Q_L B - \omega A) - 0.23 \text{ db} \} - \frac{Q_L}{\omega} \left\{ \frac{Q_L B}{Q_L B - \omega A} \right\} \ln(Q_L B - \omega A).$$

Inserting this value of the constant of integration we get the final solution for t

$$t = \frac{AQ_L}{Q_L B - \omega A} \left\{ \ln \left(\frac{a\omega W_s}{Q_L B - \omega A} \right) + 0.23 \text{ db} \right\} - \frac{Q_L}{\omega} \left\{ \frac{Q_L B}{Q_L B - \omega A} \right\} \ln \left(\frac{Q_L B - \omega A - a\omega W_s}{Q_L B - \omega A} \right). \quad (7)$$

If we define

$$\begin{aligned} 1/Q &= (B\omega L - 1/Q_L), \\ \varphi &= a\omega^2 L/Q_L, \\ W &= \varphi Q W_s, \end{aligned}$$

where $1/B\omega L$ is a negative input Q which represents the rate of energy input to the circuit, and $1/\varphi Q = W_s$ final, Eq. (7) now reduces to

$$t = \frac{Q}{\omega} \{ \ln(W) + 0.23 \text{ db} \} - \frac{Q + Q_L}{\omega} \ln(1 - W), \quad (8)$$

where the build-up curve is generated as W goes from zero to one.

Using Eq. (8), we may calculate the form of the build-up curve, and the starting time as a function of load and of noise level. Calculating first the starting time as a function of load we

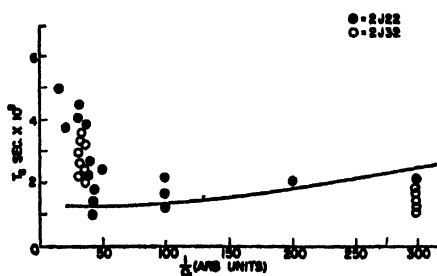


FIG. 8a. Starting time vs. load using stored energy in the magnetron alone.

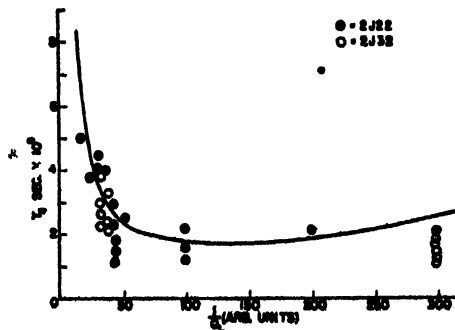


FIG. 8b. Starting time vs. load using total stored energy.

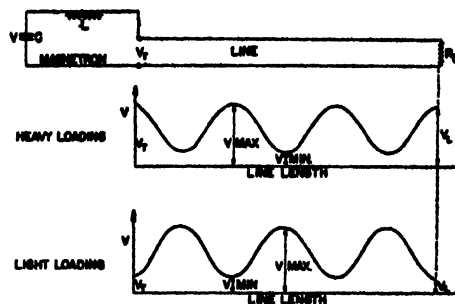


FIG. 9. Effect of load on V_T .

define the starting time as the time for W to reach the value 0.01. This is equivalent to W_s reaching 1 percent of its final value since for each load $W_s = W(W_s \text{ final})$. This definition of starting time agrees with that of More and Miley.⁵ In Fig. 8a is a plot of the starting time, as defined above, against loaded $Q(Q_L)$. The curve was calculated from Eq. (8), using a noise level of 80 db down. The open circles are data of More and Miley on 2J32 tubes, and the closed circles are data on a 2J22. It appears that the agreement is poor in the region of light loading. It would seem, therefore, that we need to examine our assumptions more carefully. We derived Eq. (8) on the assumption that the stored energy in the magnetron was the only important stored energy in the system, and that it depended only on the square of the slot voltage and not upon the load ($W_s = A V^2$).

If we consider carefully the connection of the load to the stored energy, we will find that it exerts a remarkable influence on it. In Fig. 9 are shown the standing wave patterns for heavy and light loading. It is easy to see that the voltage V_s at the back of the magnetron cavity varies widely for these two cases. This means that the stored energy in the magnetron itself varies widely for these two cases even for essentially the same power output. In addition to this, there is some energy stored in the line when the line is not matched which adds a constant amount of stored energy to either case for the same VSWR and length of line. Taking these facts into consideration, we will correct Eq. (8) for the effect of load on the stored energy.

We shall first demonstrate the importance of stored energy contributed by the load, by means

⁵ See reference 1, footnote 2.

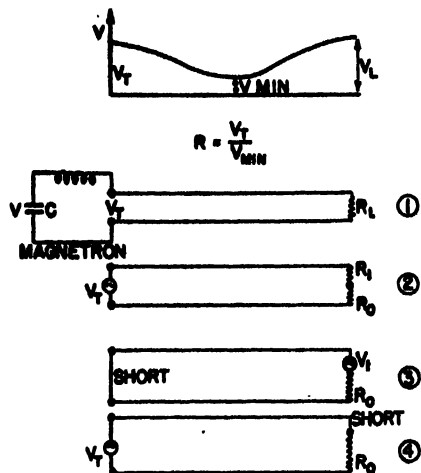


FIG. 10. Circuit for determining stored energy in line.

of the example of a shorted line section. It can be shown in general that if B is the susceptance of a resonant circuit the stored energy, E , may be written

$$E = \frac{V^2}{2} \frac{dB}{d\omega} \bigg|_{B=0} \quad (9a)$$

In the case of reactance (x) instead of susceptance we have

$$E = -\frac{i^2}{2} \frac{dx}{d\omega} \bigg|_{x=0} \quad (9b)$$

where V is the voltage at the terminals where $dB/d\omega$ is measured, and i the current at the terminals where $dx/d\omega$ is measured. This is a corollary of Foster's reactance theorem. For a coaxial line or a wave guide we know that the terminal impedance Z is defined

$$Z = Z_0 \frac{Z_L + jZ_0 \tan(\omega l/c)}{Z_0 + jZ_L \tan(\omega l/c)}$$

and the terminal admittance

$$Y = Y_0 \frac{Y_L + jY_0 \tan(\omega l/c)}{Y_0 + jY_L \tan(\omega l/c)}$$

$$Y = jY_0 \tan(\omega l/c) \quad \text{if} \quad Y_L = 0.$$

From these expressions we may calculate

$$\frac{dB}{d\omega} \bigg|_{B=0} = \left(\frac{l}{c} \right) Y_0 \sec^2(\omega l/c) \bigg|_{\omega l/c=0, \pi, 2\pi, \text{etc.}} = \left(\frac{l}{c} \right) Y_0.$$

For

$$\omega l/c = 0, \pi, 2\pi, \text{etc.} \quad l/c = h\pi/\omega,$$

where h is the number of half-wave-lengths of line. Finally

$$\frac{dB}{d\omega} \bigg|_{B=0} = \frac{h\pi Y_0}{\omega} \quad (10a)$$

Similarly, we find

$$\frac{dx}{d\omega} \bigg|_{x=0} = \frac{h\pi Z_0}{\omega} \quad (10b)$$

If we consider the line of Fig. 10, and if we define (R) as the voltage at the terminals A divided by the voltage, one-quarter wave-length down the line, we may calculate the energy stored in the line as a function of the (R) and the terminal voltage V_t .

By the theorem of compensation we may consider the load R_L replaced by a load R_0 in series with a generator of zero internal impedance which generates such a voltage V_1 , that its addition to the voltage V_0 gives the voltage drop through the original R_L . We will now consider the voltage in R_0 (matching resistance) as giving rise to energy in transit and the voltage V_1 as giving rise to stored energy in the line. First calculating the current i_1 through R_1 we will, by the theorem of superposition, replace the generator V_t by a short circuit (internal resistance of generator $V_t = \text{zero}$), then

$$i_1 = V_1/R_0,$$

since the only generator now present is the V_1 generator. Now

$$V_1 = V_t R_1 / (R_1 + R_0)$$

and

$$i_1 = \frac{V_t R_1}{R_0 (R_1 + R_0)} = (V_t / R_0) (R - 1) / R. \quad (11)$$

For $R < 1$ we use the formula

$$W_t \text{ (stored energy in the line)} = -\frac{i^2}{2} \frac{dx}{d\omega} \bigg|_{x=0}$$

and obtain

$$W_i = \frac{V_i^2}{R_0^2} \left(\frac{R-1}{R} \right)^2 (\hbar\pi R_0/2\omega).$$

For $R > 1$ we use formula (9a) giving

$$W_i = V_i^2 \left(\frac{R-1}{R} \right)^2 (\hbar\pi G_0/2\omega),$$

which we see to be the same as for $R < 1$ giving the stored energy in the line for any R

$$W_i = V_i^2 \left(\frac{R-1}{R} \right)^2 (\hbar\pi G_0/2\omega). \quad (12)$$

The energy stored in the magnetron (W_m) may be calculated from the transmitted $Q(Q_i)$.

$$Q_i = \frac{2\pi \text{ energy stored}}{\text{energy lost/cycle}} = \frac{2\pi W_m}{V_i^2 G_L/f} \quad (13)$$

or

$$W_m = V_i^2 Q_i G_L/\omega.$$

Combining (12) and (13) to get the total stored energy, W_s , we have

$$W_s = W_i + W_m = \frac{V_i^2 G_0}{\omega} \left\{ R Q_i + \frac{\hbar\pi}{2} \left(\frac{R-1}{R} \right)^2 \right\}. \quad (14)$$

In order to consider the variation of A , we must transfer the voltage V_i back to the slots of the magnetron. To do this we will use Eq. (1)

$$V^2 = P_0 Q_i s_0. \quad (1)$$

Now the power output is

$$P_0 = V_i^2 G_L = V_i^2 R G_0$$

for an even number of quarter-wave-lengths of line between the load and the terminals where V_i is measured.

For the magnetron $s_0 = (L/C)^{\frac{1}{2}}$ enabling us to transform Eq. (1) into

$$V^2 = V_i^2 R G_0 Q_i (L/C)^{\frac{1}{2}}$$

and the total stored energy, W_s , becomes

$$W_s = \frac{V^2 s_0}{\omega} \left\{ 1 + \frac{\hbar\pi}{2 Q_i^*} \left(\frac{R-1}{R} \right)^2 \right\}, \quad (15)$$

where $g_0 = 1/s_0 = (C/L)^{\frac{1}{2}}$ for the magnetron, and Q_i^* is the transmitted Q for matched line. Since

by definition $W_s = A V^2$ we now see that A depends on the load, here represented by R .

$$A = g_0/\omega \left\{ 1 + \frac{\hbar\pi}{2 Q_i^*} \left(\frac{R-1}{R} \right)^2 \right\}. \quad (16)$$

Examining Eq. (16) we see that for matched line ($R=1$). $A = g_0/\omega = 1/\omega^2 L$, as before. For $R > 1$ (heavy loading) the term

$$(\hbar\pi/2 Q_i^*) \left(\frac{R-1}{R} \right)^2$$

approaches a constant finite value as R increases. Thus, the total stored energy approaches a finite value as the loads get heavier. Now as R decreases on the side of light loading, the term

$$(\hbar\pi/2 Q_i^*) \left(\frac{R-1}{R} \right)^2$$

goes to infinity, so that the total stored energy for the case of light loading approaches infinity as R goes to zero. This is quite reasonable since V_i must remain finite even when the $VSWR$ goes to infinity, so that when V_i is a maximum of the $VSWR$, the total stored energy remains finite, but when V_i is a minimum of $VSWR$ the

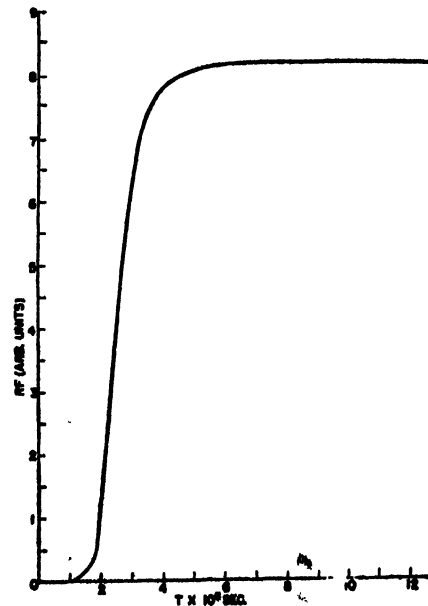


FIG. 11. Calculated build-up curve.

total stored energy should go to infinity as the (R) goes to zero.

Similarly, the loaded $Q(Q_L)$ used in Eq. (5) depends on the total stored energy,

$$Q_L = \frac{2\pi W_s}{(P_0 + \text{loss})/f} = \omega W_s \epsilon / P_0, \quad (17)$$

where ϵ is the circuit efficiency. Using (16) and (17) in Eq. (7) we may calculate the starting time as a function of load with the stored energy in the line fully accounted for. The results of this computation are shown in Fig. 8b. Here the various constants used are $h=8$ half-wave-lengths (a reasonable distance for a standing wave detector and a tuner) noise level -80 db down, $Q_L^* = 100$. Here it is seen that the agreement between the calculated curves and the results of More and Miley (circles) is quite good—in the regions of both light and heavy loading.

SECTION III

Results and Conclusions

In Fig. 11 is given a curve of stored energy against time as calculated from Eq. (8). This curve may be compared with the solid curve of Fig. 4, for it is a build-up curve of the magnetron. It also refers to zero time of rise of the voltage pulse. The agreement is seen to be quite good.

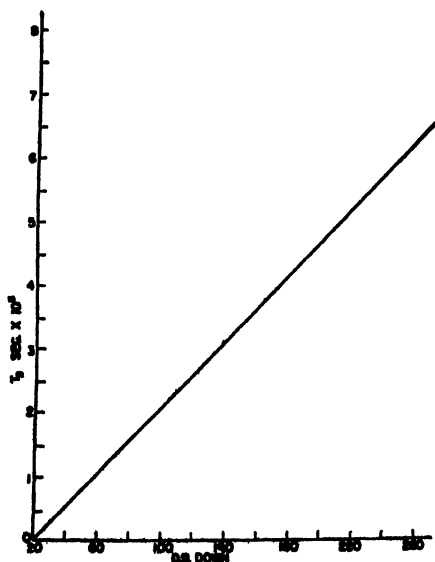


FIG. 12. Variation of starting time with noise level.

The slight difference of the top bends in the curves may be ascribed to the fact that the curve of Fig. 4 was corrected from experimental curves taken with voltage pulses that oscillated slightly upon reaching their peak values. The form of the build-up curve for different loads can be calculated by using Eq. (7) corrected by (16) and (17) but it does not seem to be worth while here.

In Fig. 12 is shown a curve of starting time against noise level for a matched load. This curve also was calculated from Eq. (8). It is clear that the noise level does not affect the starting time very much for any values which are physically possible. Noise levels from 160 db to 80 db down all give starting times of the right order of magnitude.

In this connection it was suggested by Dr. Altar that an undesirable mode could be suppressed by injecting an r-f voltage of the frequency of the desired mode, thus giving it a head start, so to speak, over the undesired mode. Such an experiment was performed by Dr. Colman of this laboratory and was found to be successful.

In conclusion several points should be emphasized. First, it appears from the available data that a 2J32 magnetron has a limiting load beyond which it will not oscillate. This load appears to be in the neighborhood of a loaded Q of 15. At such a load the starting time becomes infinite so that the experiment of More and Miley should show a sharp increase of starting time if the curves were pursued to very low loaded Q values. Dr. Colman has made magnetrons in the 3-cm band which were loaded to a Q_L of 12 with a frequency insensitive load, and found it quite impossible to get them to oscillate in the mode so loaded. The same tubes did oscillate when loaded to a Q_L of 20, showing that for a Q_L in the neighborhood of 15 oscillation ceased. This piece of evidence is cited in support of the result given above. Such evidence establishes with certainty the fact that the curve of Fig. 6 approaches zero with a finite slope.

Second, it should be pointed out that the curves of Fig. 6 and Fig. 7 are plotted from data taken with the d.c. voltage on the magnetron held constant, not the input power. This means that even though the power generated is a

monotonically increasing function of the r-f voltage on the slots, the conversion efficiency is not. In Fig. 13 is given the variation of conversion efficiency with the ratio of the r-f voltage on the slots of the magnetron to the applied d.c. voltage. It is seen that there is a maximum of conversion efficiency when $V_{r-t}/V_{d.c.} = 0.65$. This result is in good agreement with the calculations of Dr. Altar⁶ who gets such a maximum at $V_{r-t}/V_{d.c.} = 0.55$, in his analysis of a bunching and catching mechanism as applied to energy conversion in magnetrons. The fact that the ratio obtained here is slightly larger than that of Altar is explained by the fact that the ratio of Altar is computed at a slight distance from the anode surface. This means that both the V_{r-t} and the $V_{d.c.}$ are somewhat reduced, but since the V_{r-t} falls off faster as one leaves the anode surface, the ratio should be somewhat less.

The variation of starting time with load, as shown in Fig. 8b, would appear to lend itself readily for use in "mode" suppression in magnetrons. The mode desired could be loaded in such a manner as to keep its starting time low while the same physical arrangement could be designed to load the undesirable mode in such a manner that it would have a much longer starting time.

⁶ William Altar, private communication which will be published shortly.

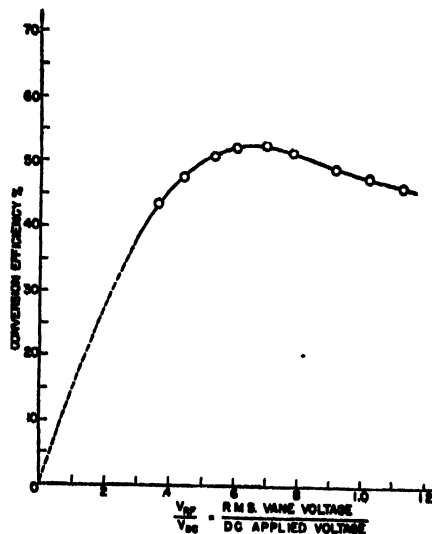


FIG. 13. Conversion efficiency vs. the ratio of slot to d.c. voltage.

This is possible because a line length can be chosen for two modes of different frequency so that one will have a standing wave pattern in the line similar to that shown in Fig. 9 for "heavy loading" and the other will have a pattern similar to that shown for "light loading."

The author wishes to acknowledge the assistance of Dr. John W. Coltman without whose constant advice this paper would not have been completed.

The Measured Impedance of Cylindrical Dipoles

D. D. KING

Craft Laboratory and Research Laboratory of Physics, Harvard University, Cambridge, Massachusetts

(Received April 2, 1946)

The input impedance of cylindrical dipoles is considered from the experimental point of view. Errors in measuring methods are considered and the results of a recent determination of dipole impedance are displayed.

INTRODUCTION

IN recent years considerable advance has been made in the theory of antennas of various shapes. The theoretical treatments regularly include an approximation of the cylindrical dipole. The purpose of this paper is to provide experimental verification of the results obtained by mathematical solutions of the problem.

The significance of experimental results depends upon an adequate definition of the quantity being measured as well as upon the precision of the experimental technique. Differences between theoretical and physical boundary conditions are considered briefly in this connection. The experimental methods are described and their accuracy is specified. With this information the results are useful in the interpretation of other data on antenna input impedance. The measured values displayed in this paper have already been quoted to this end elsewhere.^{1,2}

PHYSICAL BOUNDARY CONDITIONS

Analytical solutions for the cylindrical dipole must satisfy Maxwell's equations and the pre-

scribed boundary conditions. In general certain simplifications in the boundary conditions are required to obtain a mathematical solution to this problem.³ The discrepancy between theory and experiment attributable to differences in the boundary values may be minimized by careful experimental methods as well as by refinements in analysis. The dipole model used for the present measurements represents certain boundary values which are now considered.

The antenna configuration consisted of a half-dipole and image plane. For measuring purposes the dipole is mounted on the inner conductor of a concentric line the outer conductor of which is attached to the image plane. The coaxial line-spacing produces an annular gap at the center of the image plane. The antenna is driven by the fields in this region and hence the input impedance is also defined here. The currents at the driving point are indicated in Fig. 1.

There would seem to be a phase lag between the actual currents on the image plane and those which would exist in the ideal case of a base-driven dipole. The amount of lag is $\beta(b-a)$, where a and b are the inner and outer radii of the line and $\beta = 2\pi/\lambda$. The magnitude of the quantity $\beta(b-a)$ is determined by other considerations to be examined later. Thus, a principal consequence of the concentric line-spacing is to introduce a lag in the currents in the image antenna proportional to this quantity. Discontinuities at the junction of line and antenna conductors may cause an added lumped capacitance. The geometry at the junction of line and antenna shown in Figs. 2a and 2c applies to the present measurements. Arrangements shown in Fig. 2b were not used.

Another consequence of the finite spacing

³ L. Brillouin, *Elec. Commun.* 21, 4 and 22 (1943).

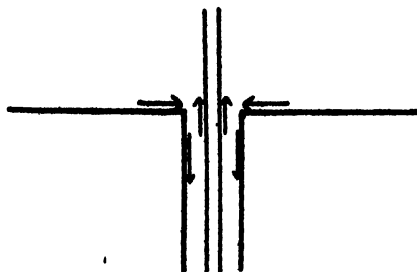


FIG. 1. Currents at the driving point of a half-dipole and image plane.

¹ R. King and D. Middleton, *Quart. App. Math.* 3 (Jan. 1946).

² D. Middleton and R. King, *J. App. Phys.* 17, 273 (1946).

$(b-a)$ is to distribute the driving field along a segment of the center conductor of approximate length $(b-a)$. The physical boundary condition for concentric-line excitation therefore does not contain well-defined terminals; instead it involves a potential difference distributed about the base of the half-dipole. Thus the actual generator to be approximated by theoretical analysis is in the form of a thick circular ring, or toroid, which maintains a potential difference along a continuous segment of conductor.

Impedance measured between terminals with appreciable spacing is of course not precisely defined. The fundamental relation for scalar and vector potentials in the interval AB is the following

$$\int_A^B \mathbf{E} \cdot d\mathbf{s} = - \int_A^B \left\{ j\omega \mathbf{A} + \frac{d\phi}{ds} \right\} ds. \quad (1)$$

In case the interval between the terminals A and B is very short, the term in the vector potential $\mathbf{A}$ may be neglected. The integral then becomes independent of the path AB . The usual definition of impedance in terms of the scalar potential difference may then be formulated

$$Z_{AB} = (\phi_A - \phi_B) / I. \quad (2)$$

Evidently a slight error is always present at the driving point due to the finite line-spacing. In an exact comparison with theory this relatively minor source of error may be properly considered. Factors that prevent an indefinite reduction of the spacing at the terminals are examined in the next section.

CHOICE OF DIMENSIONS

The scale of the apparatus and antennas used depends on the choice of wave-length. In making this choice, two basic limitations are encountered. At long wave-lengths the earth introduces errors since it is an imperfect conductor and not well suited for use as an image plane. On the other hand, extremely short wave-lengths decrease the inherent accuracy of the apparatus. This is because dimensions and tolerances decrease in proportion to the wave-length. For operation in the laboratory the range between 10 cm and 100 cm is indicated. In this region an accuracy of 0.1 mm in position is adequate and readily

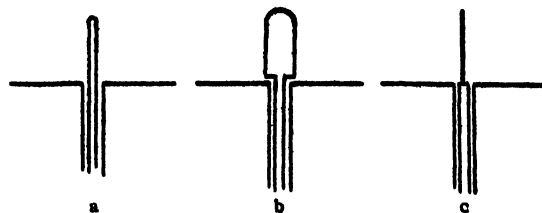


FIG. 2. Half-dipole and coaxial line configuration.

attainable, as are the usual construction tolerances of ± 0.001 inch. An image surface of ample proportions may also be provided without resort to a field station.

The choice of characteristic resistance of the measuring line is dictated by the loads to be measured. In general, the best accuracy is achieved when power is efficiently transferred to the load. However, since impedance measurement often involves the use of cables, the available fittings and cables must be considered. In this instance a line resistance of 49 ohms was chosen.

Having thus fixed the ratio of conductor diameters, the magnitude of the quantity $(b-a)$ at the image surface is fixed by the antenna diameters to be used. This precludes very small values of $(b-a)$. The small dimensions of the line at this point are not favorable to measuring procedures since closer tolerances are required in a thin line. Accordingly, a 16-inch linear taper was used in the present measurements which everywhere maintained the characteristic resistance. No reflections from this tapered section were detectable and the measuring line diameter could thereby be set at 1 inch while the hole in the image plane was only $\frac{1}{16}$ inch, which gives $\beta(b-a) = 0.05$ rad. at the standard wave-length of 28.00 cm. The phase lag introduced by the hole in the image plane, though small, is therefore not entirely negligible.

MEASURING METHODS

In making antenna impedance measurements it is possible to choose one of two common parameters. Most theoretical treatments of the subject use the ratio of antenna length to diameter as the parameter. This is commonly written in the form of $\Omega = \log(2h/a)$ where h is the half-length and a the radius of the antenna. From the experimental point of view this implies



FIG. 3. Measuring apparatus: the line is at the left; monitors, amplifiers, and oscillator at the right.

varying the frequency. In order to include all lengths from the half-wave-length down to very short antennas, a very wide range of frequencies must be used. This is troublesome since the errors introduced by the finite image plane will change. Moreover, the tuning of the measuring equipment and the detector calibration all require attention as the frequency is varied over a wide range. Accordingly, another parameter was chosen in which a/λ is held constant and the half-length h changed. Since the theoretical parameter Ω is logarithmic, and hence slowly varying, the discrepancy between theory and experiment is negligible in any small interval. By measuring various thicknesses a curve with constant Ω can be obtained. The experimental procedure involves shortening the antenna for each reading. This is hardly more time-consuming than the readjustment of the apparatus would be, and permits greater accuracy in the results for the reason mentioned.

In order to assure maximum accuracy it seemed advisable not to rely exclusively on the

standard standing-wave-ratio (SWR) method of measurement. For small losses in line and load this method is open to criticism because of the possibility of detector loading. An apparatus permitting both the conventional SWR method and the resonance curve method to be used provides greater flexibility in obtaining results for all values of terminal impedance. The use of resonance curves implies a movable piston at one end of the line, and this eliminates the possibility of using solid dielectric filled lines. The presence of dielectric beads, strips, or pins in the line is highly undesirable because of the discontinuities and reflections occasioned by them. It is possible to calculate the position of two beads with perfect dielectric which will produce no reflections. However, the calculation in general demands that the terminal impedance be known.

In order to avoid the errors due to movable beads and other types of dielectric supports, the measuring line used for the present measurements was designed to avoid the use of all supports. This was accomplished by mounting the conductors vertically. Several important advantages accrue from this type of mounting. Thus, the image plane could be located several feet above the operator, thereby minimizing disturbing effects from objects in the room. The weight of the inner conductor and attached antenna is borne by the brass base-plate at the bottom of the line. At the extreme upper end of the line a thin, perforated polystyrene wafer is inserted flush with the surface of the image plane. This wafer, being less than 0.004λ thick (1 mm), may be treated as a lumped capacitance in parallel with the load. The reactance introduced by the polystyrene may be measured and removed from the final results by subtracting its admittance from the load admittance. The fringing, or throat effect, at the open end of the line is probably also somewhat compensated by this procedure.

The wafer support maintains the line spacing well since gravitational effects are eliminated. The space between the two coaxial conductors is then left free to permit a piston to travel up and down for resonance curve measurements. Adequate rigidity is obtained from the silvered ground steel (drill rod) inner conductor and the $\frac{1}{4}$ -inch wall of the aluminum outer conductor.

Tolerances were better than ± 0.001 inch. The radii of the inner and outer conductors were $a = 0.5588$ cm and $b = 1.2738$ cm. The calculated characteristic resistance is $R_0 = 49.38$ ohms; the attenuation constant $\alpha = 3.78 \times 10^{-4}$ neper/cm. At the standard wave-length of 28.00 cm, the phase constant was $\beta = 0.2244$ radian/cm.

A movable tuned probe is provided for detecting the voltage amplitude along the line; probe depth is readily adjustable by a lead screw. The output of a Sylvania 1N21B crystal in the probe is fed through a selective amplifier into a Ballantine voltmeter. The voltage gain of this amplifier is 78 db and the noise level about 78 db below one millivolt. Extensive grounding proved to be essential in maintaining the noise level. Power is furnished by a 2C40 triode in a tuned cavity; 1100-cycle square-wave modulation is applied. The measuring line and associated equipment is shown in operating condition in Fig. 3.

MEASURING PROCEDURE

For low values of damping the resonance-curve method of measurement was used; for high damping, the standing wave ratio supplemented the resonance curve data. Circuit adjustments to permit changing from one method to another could be made quickly. Particular emphasis was placed upon the monitoring in order to eliminate fluctuation errors.

In resonance curve measurements it is essential that the oscillator be loosely coupled to the line. This is accomplished by the circuit shown in Fig. 4. A concentric-line-type wave meter and a bolometer amplitude-monitor are continuously coupled to the circuit. Constant information on the output amplitude and frequency is furnished by direct reading meters. Amplifiers provide a

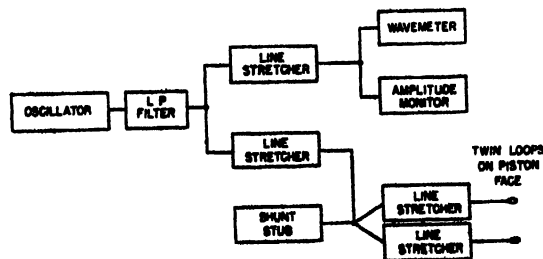


FIG. 4. Block diagram of the high frequency circuit.

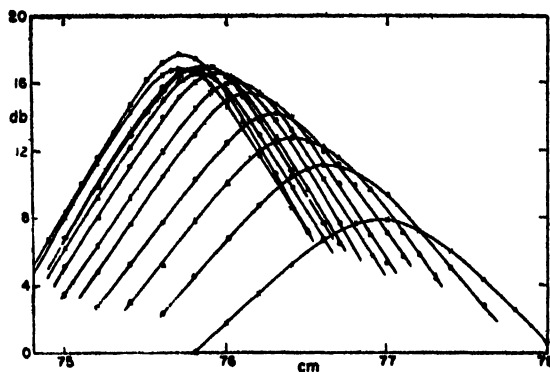


FIG. 5. Original data in the form of resonance curves. The wave-length is 28 cm and the width was taken at a level 3.16 db below the peak of each curve.

monitor sensitivity equal to the sensitivity of the detector on the line. If the amplitude- and frequency-monitors are undisturbed by the tuning of the measuring line through the required interval, then the coupling between oscillator and measuring line is loose, *by definition*.

In order to provide flexibility in adjustment two separate coaxial feeders are provided in the piston assembly. Each feeder terminates in a loop 3 cm long on the face of the upper piston. Tandem pistons with 3-cm brass flutes assure complete decoupling between the line and the cavity located between the back of the pistons and the line base-plate. Each feeder cable is attached to an independent phasing stub in a twin line-stretcher. A shunt stub is placed to form a four-way junction with these feeders and the line from the oscillator. This shunt stub forms the mutual impedance Z_{12} between the oscillator (primary) and measuring line (secondary). Suitable stops on the stub permit rapid adjustment of the coupling between oscillator and line.

For standing-wave-ratio measurements a larger amplitude is required along the line, since the minima of the distribution curve must be measured. The resonance curve procedure deals only with the peak of the resonance curve and hence requires less power input to obtain adequate deflections. Any desired degree of coupling can be obtained readily by means of the shunt stub mentioned above. Provision is made to permit accurate reproduction of the settings of this stub. This is needed in the resonance curve procedure, which requires the

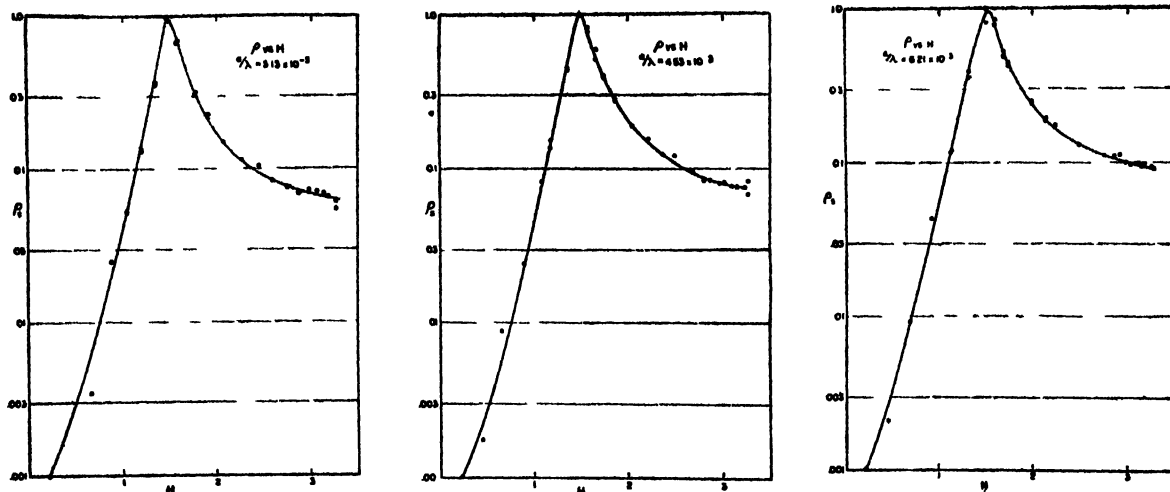


FIG. 6. Termination damping ρ in nepers as a function of $H = \beta h$, the dipole half-length in radians; the radius of the dipole is α . Solid points are from resonance data; circles from standing-wave data.

terminal impedance at the sending end to remain constant between comparisons with the standard termination.

An auxiliary probe near the sending end of the slotted section of line may be inserted to permit a direct test of probe loading. Probe effects cancel out in the resonance curve procedure, but the standing-wave method may suffer from probe loading effects at higher standing-wave ratios.⁴ No loading effects could be detected for the standing-wave ratios measured.

ACCURACY OF MEASUREMENTS

The accuracy of which the apparatus is capable depends principally on two factors: first, the precision of linear measure along the line; and second, the accuracy of the detector reading. Steel scales and verniers permit readings on the present apparatus to within 0.1 mm. Since a sliding contact is involved, some irregularities might be expected. However, the original data yielded uniformly smooth resonance curves. A sample family of resonance curves for an antenna load is shown in Fig. 5. A fair estimate of the accuracy of the data on the position of the curves is probably better than ± 0.1 mm or $\pm 0.0004\lambda$. This is because the aggregate of all the points determines this position, which is therefore more accurately known than the location of any one point. Tests on the standard copper-plate termination made at regular inter-

vals bear this out. The width of the curves is given to nearly the same precision, but additional error is introduced here in fixing the level at which the width is to be measured.

The width of the curves was measured at $(1/\sqrt{2})V_{\max}$. This level, somewhat higher than the usual half-power point, was chosen to extend the range of the method in the direction of high damping (low SWR), thereby securing a greater overlap with the standing-wave-ratio method. Increased accuracy in measuring the width of the curves is required at the level chosen, but the limit of error mentioned above is sufficiently small to permit this. The square of the amplitude level was read on a Ballantine voltmeter. An accuracy within ± 0.1 db was assumed. Repeated calibration tests were made and the deviations on the calibration curve confirm this as the likely error.

The original experimental results are in terms of termination phase shift Φ and termination damping ρ .^{5,6} The resistance and reactance of the load are functions of both these independent variables.

$$R = 2R_0 \frac{\sinh 2\rho}{\cosh 2\rho - \cos 2\Phi}, \quad (3)$$

$$X = -2R_0 \frac{\sin 2\Phi}{\cosh 2\rho - \cos 2\Phi}. \quad (4)$$

⁵ R. King, J. App. Phys. 14, 577 (1943).

⁶ R. King, Proc. I.R.E. 29, 640 (1941).

⁴ Altar, Marshall, and Hunter, Proc. I.R.E. 34, 33 (1946).

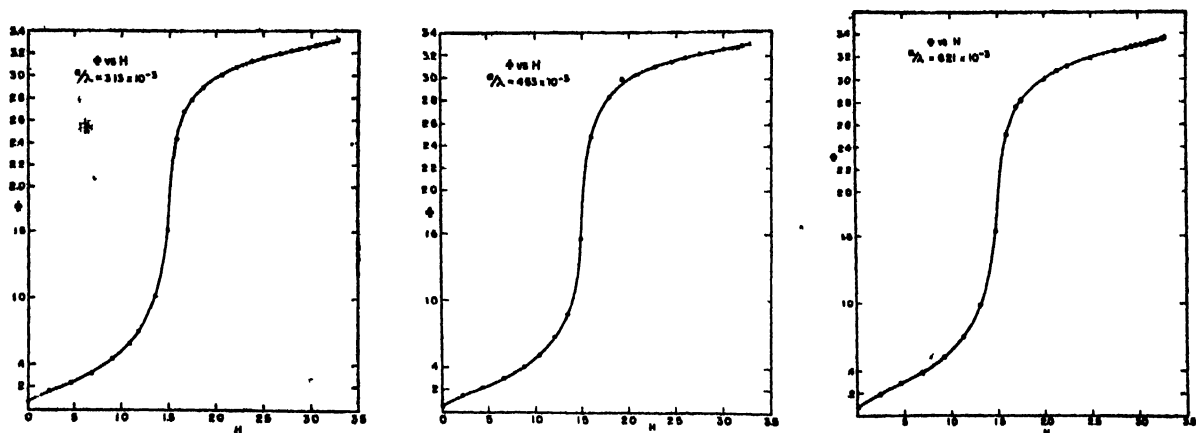


FIG. 7. Termination phase shift in radians as a function of the dipole half-length in radians.

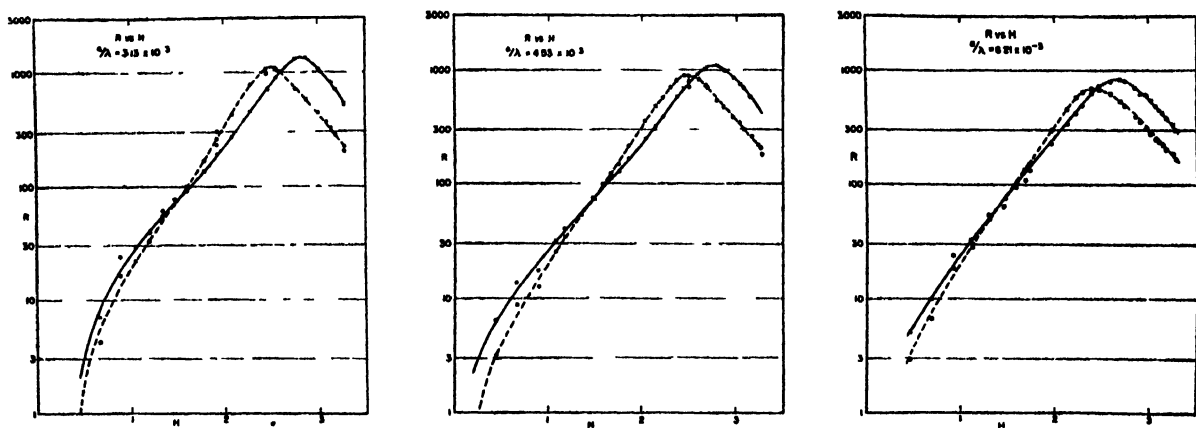


FIG. 8. Input resistance of dipole of radius a as a function of its half-length in radians. The solid curve has been compensated for the shunting effect of the polystyrene support; the dashed curve has not.

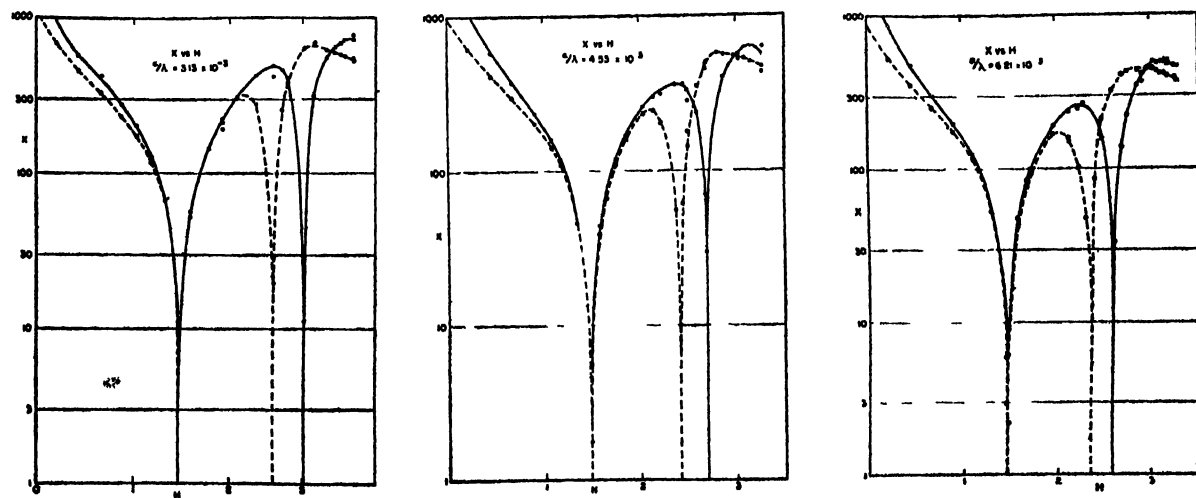


FIG. 9. Input reactance of a dipole of radius a as a function of its half-length in radians. The solid curve has been compensated for the shunting effect of the polystyrene support; the dashed curve has not. The sign is not contained in the logarithmic plot.

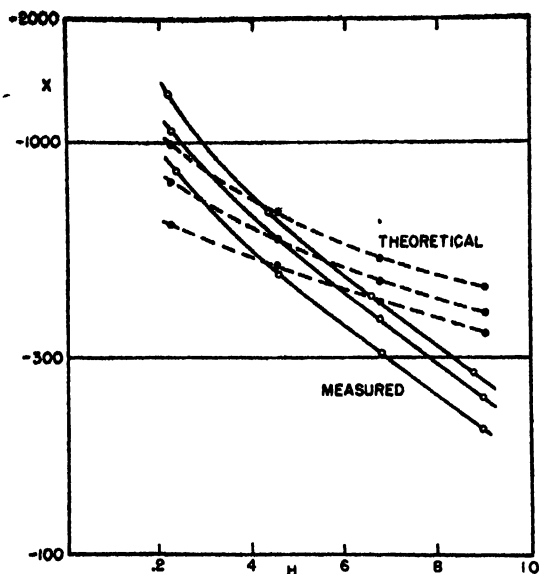


FIG. 10. Reactance of short dipoles. The measured curves are taken from Fig. 9 and the theoretical values calculated electrostatically.

The factor 2 in Eqs. (3) and (4) converts the results to symmetrical dipole impedance.

The accuracy of the ohmic results is not so easily stated, since the transformation from Φ and ρ to R and X may increase or reduce the inaccuracies in the measured variables. A measurement depending principally on the phase-shift function Φ was made at regular intervals to determine the impedance of the open end with associated wafer support. A value of $-j1440$ ohms ± 5 percent was obtained. Since polystyrene is somewhat affected by the humidity, some of the deviation recorded is probably not caused by the measuring technique.

Measurements made at large terminal damping depend principally on the damping function ρ . Independent determinations of ρ were made, using the resonance-curve method and the standing-wave-ratio method. The two methods agreed very closely. The experimental points from both methods are plotted in the curves. An error of the order of 5 percent is indicated by the location of the points along the curves for ρ . Values taken from the smooth curve should be more accurate. A much smaller deviation of the experimental points is apparent in the curves of the phase-shift function Φ . The curves are shown in Figs. 6 and 7.

The curves of R and X shown in Figs. 8 and 9 show both compensated and uncompensated values. The compensation involves elimination of the parallel reactance of -1440 ohms introduced by the polystyrene wafer support. Evidently the presence of even a small section of dielectric influences the results appreciably. The significance of conditions at the driving point is therefore borne out by the experimental results. The curves for ρ and Φ are slowly varying when the corresponding R and X curves exhibit rapid changes. Therefore the use of the ρ and Φ plots effectively increases the number of experimental points available in critical regions of the resistance and reactance graphs. Several additional points were chosen in this manner about the resonant and anti-resonant lengths.

The image plane used consisted of a square sheet of aluminum 90 cm on a side, supplemented by eight brass rods interconnected by heavy wire. The total area of this assembly was about eight times that of the center section. Data taken with only the inner section in place differed from values taken with the added section by less than 3 percent for several antenna lengths. The phase function Φ remained unaltered, while the damping ρ increased slightly with the size of the plane. A 5 percent increase in

TABLE I. Critical values of cylindrical antenna, $\Omega = 10$.

	Experimental D. D. King	King-Middleton (second order)
$(R_0)_{\text{anti-res}}$	820	860
$\pi - \beta h_{\text{anti-res}}$	0.60	0.614
$ X _{\text{min}}/ X _{\text{max}}$	1.95	1.8
$(R_0)_{\text{res}}$	71.5	71
$\pi/2 - \beta h_{\text{res}}$	0.098	0.094
$(R_0)_{\beta h = \pi/2}$	85	88
$(X_0)_{\beta h = \pi/2}$	47	42.5

TABLE II.

	D. D. King compensated	D. D. King uncompensated	Brown and Woodward	$a/\lambda \times 10^{-3}$
$(R_0)_{\text{anti-res}}$	1320	1080	1340	3.13
	1080	880	1050	4.53
	820	700	800	6.21
$\pi - \beta h_{\text{anti-res}}$	0.36	0.70	0.47	3.13
	0.44	0.72	0.59	4.53
	0.60	0.79	0.68	6.21
$ X _{\text{min}}/ X _{\text{max}}$	1.55	2.0	1.8	3.13
	1.88	2.32	2.3	4.53
	1.95	2.57	4.2	6.21

damping was recorded for flat-topped antennas as compared with round-topped types of equal length. Again the phase function Φ remained constant. The data available at present are inadequate to permit conclusions from these special results pertaining to the image plane and the end-surfaces. Further measurements are indicated to elucidate the effect of the end-surfaces and of the image plane.

COMPARISON OF RESULTS WITH THEORY AND WITH OTHER MEASUREMENTS

Several critical points in the impedance characteristics of an antenna may be chosen for comparison of the experimental results obtained with calculated values. In general, the maximum, minimum, and zero values of reactance are determined primarily by phase-shift data, while the resistance is also a function of the damping. In view of the greater precision of phase-shift determinations, somewhat greater accuracy may be expected in the reactance data than in the resistance values. Table I shows the results of a recent theoretical treatment and of the present measurements. The theoretical aspects of the numerical results have been discussed with a previous publication of this table.³

A plot of certain critical values of resistance

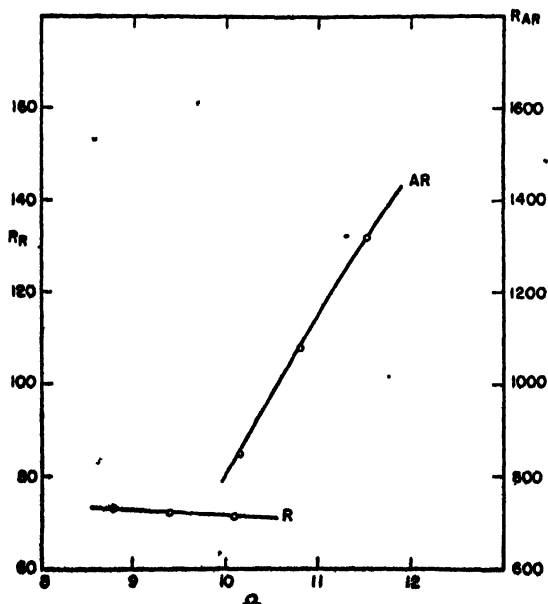


FIG. 11. Resonant and anti-resonant resistance as a function of the thickness parameter $\Omega = 2 \ln(2h/a)$.

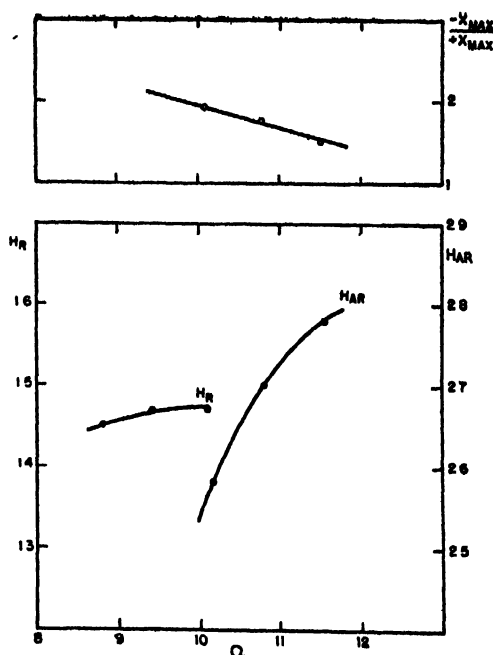


FIG. 12. Top: ratio of the magnitude of maximum negative and positive reactance; bottom: resonant and anti-resonant half-length in radians, all as a function of the thickness parameter $\Omega = 2 \ln(2h/a)$.

and reactance is shown in Figs. 10, 11, and 12. The curves of the reactance of very short dipoles are of particular interest. In case the antenna is short ($\beta h = H < 1$), an electrostatic calculation of the capacitance can be made⁷ and X set equal to $-1/\omega C$. A very reliable theoretical result is thereby obtained. Agreement between theory and experiment here appears good. In case of extremely short antennas, the measurement is no longer accurate since the hole in the image plane and the antenna length are of the same order of magnitude. Likewise, antennas longer than $H = 0.5$ radian no longer admit use of a lumped capacitance determined from an electrostatic analysis.

A comparison with experimental results of Brown and Woodward is included in Table II. These writers used a 5-meter wave-length with a conventional slotted line. The agreement is rather good, with the exception of the reactance ratio. This is to be expected, however, since Brown and Woodward⁸ directed their measure-

⁷ R. King and C. W. Harrison, *J. App. Phys.* 15, 170 (1944).

⁸ G. H. Brown and O. M. Woodward, *Proc. I.R.E.* 33, 257 (1945).

ments toward cases of practical engineering significance. Considerable base capacitance is encountered in base-driven dipoles attached through standard cable connectors. The effect of base capacitance is clearly indicated by the uncompensated results of the present measurements.

CONCLUSION

Graphs of the measured impedance characteristics of cylindrical dipoles have been presented. Separate plots for various thicknesses are intended to display the individual experimental points clearly. The results are given in this fashion together with specifications of the

experimental procedure and accuracy in order to facilitate the correlation between antenna theory and experiment. Although the available apparatus cannot conform in full to the theoretical models described, a close approximation was generally achieved.

The writer is indebted to Professor R. W. P. King for many valuable suggestions in connection with the measurements described. These results are incorporated in a portion of the writer's Ph.D. thesis at Harvard University. The apparatus involved was constructed at Central Communications Research, Cruft Laboratory, under the auspices of the Office of Scientific Research and Development.

Elementary Treatment of Longitudinal Debunching in a Velocity Modulation System

EUGENE FEENBERG

Washington Square College, New York University, New York, New York

(Received May 14, 1946)

Certain configurations of anode, grid sections, and uniformly emitting cathode produce rectilinear motion of the electrons, both with and without velocity modulation. These systems are (a) plane parallel electrodes and grid sections of unlimited extent, (b) coaxial circular electrodes and grid sections of infinite length along the axis, and (c) concentric spherical electrodes and grid sections. The uni-dimensional character of the motion permits a complete evaluation of the effect of space charge on the bunching process in the range of drift space where overtaking (crossing of orbits) does not occur.

THE theory of velocity modulation has been developed from two distinct points of view exemplified in D. L. Webster's kinematic analysis of the bunching process¹ and Hahn, Metcalf, and Ramo's treatment of the electron beam as a medium for the propagation of electromagnetic disturbances.² There exist problems in which both points of view can be combined to advantage, retaining the rigorous foundation of the second and also a large measure of the practical usefulness of the first. It is the purpose of the current note to present a unified treatment of three closely related problems of this type.

Certain configurations of anode, grid sections, and uniformly emitting cathode produce recti-

linear motion of the electrons, both with and without velocity modulation. These systems are (a) plane parallel electrodes and grid sections of unlimited extent,³ (b) coaxial circular electrodes and grid sections of infinite length along the axis, and (c) concentric spherical electrodes and grid sections.

The uni-dimensional character of the motion permits a complete evaluation of the effect of space charge on the bunching process, subject to the following assumptions and limitations:

(1) In the space between buncher and catcher grid sections the negative space charge produced by the steady electron current is exactly neutralized by a positive ion space charge. In order to secure a precise mathematical problem it may

¹ D. L. Webster, *J. App. Phys.* 10, 501 (1939); R. H. Varian and S. F. Varian, *J. App. Phys.* 10, 321 (1939).

² W. C. Hahn and G. F. Metcalf, *Proc. I.R.E.* 27, 106 (1939); W. Ramo, *ibid.* 27, 757 (1939); W. C. Hahn, *Gen. Elec. Rev.* 42, 288 (1939), 42, 497 (1930).

³ Space charge effects in system *a* are discussed in reference 1 and also in unpublished notes by W. W. Hansen.

be assumed that the positive ions move with the same velocity as the incoming electrons.³ There is then no steady field (neither electrostatic nor magnetostatic) acting upon the beam.

(2) The ratio of beam velocity to the velocity of light is a small quantity ($v_0/c \ll 1$).

(3) The ratio of buncher voltage to beam voltage is a small quantity ($\alpha = V_1/V_0 \ll 1$). Here the form $V_1 \sin \omega t_1$ is taken for the instantaneous voltage across the buncher grids.

Velocity modulation at the buncher causes hf deviations from the steady-state electron distribution. These deviations produce a hf component of electric field parallel to the direction of propagation. The field in turn modifies the bunching process. We seek a self-consistent solution of the system of field equations and equations of motion in the sense that the field together with the velocity modulation at the buncher produces exactly the motion required to give rise to the field. In the theory of atomic structure problems of this general character are solved by methods of successive approximation (Hartree, Hartree-Fock equations). Fortunately a simple analytic solution exists in the present case.

Let z represent the distance measured along the direction of propagation (with the origin at the axis or center of symmetry in systems b and c). In the usual notation $\tau = \omega t_1$ is the phase at which an electron passes the effective center z of the buncher grid space and $\theta + \omega(z - z_1)/v_0 = \omega t$ is the phase at which the same electron reaches the position z . Suitable independent variables are provided by z and t or by z and θ . Thus any hf quantity can be expressed as a function, $f(z, \theta)$, of z and θ or as a function, $f'(z, t) \equiv f(z, \theta)$, of z and t .

The boundary condition which correlates the field in the drift space with the amount of velocity modulation is furnished by the equation

$$\dot{z} = v_0 [1 + \frac{1}{2} \mu \alpha \sin \tau], \quad z = z_1. \quad (1)$$

Here v_0 is the beam velocity, μ the beam coupling coefficient,⁴ and α the ratio of buncher voltage to beam voltage.

⁴ For system a

$$\mu = \sin(\frac{1}{2}s)/\frac{1}{2}s$$

in terms of the transit angle s through the buncher grid space. The same relation holds for systems b and c if the buncher grids are closely spaced in comparison with the mean distance of the grids from the axis or center of symmetry.

A point of departure for the calculation is supplied by the tentative hypothesis that the relation between θ and τ is formally the same as in the kinematic bunching theory. Now the bunching parameter appears as an unknown function of z in the cycloidal equation¹

$$\theta = \tau - X(z) \sin \tau. \quad (2)$$

The condition $X(z) \leq 1$ holds throughout the following discussion. This restriction is necessary and sufficient to exclude crossing of orbits and allows the hf component of current density to be computed from the relation

$$i(z, \theta) = i_0 \left(\left(\frac{d\tau}{d\theta} \right) - 1 \right) \left(\frac{z_1}{z} \right)^n, \quad (3)$$

in which i_0 is the entering current density at the center of the buncher grid section. The integer n has the values 0, 1, and 2 for systems a , b , and c , respectively. Equation (3) expresses the conservation of charge if no overtaking (or crossing of orbits) occurs in the beam. The additional factor $(z_1/z)^n$ corrects for the expansion of the beam with distance from the cathode. According to Eq. (2) $\tau = 0$ implies $\theta = 0$ or $t = (z - z_1)/v_0$. Thus electrons at the center of the bunch move uniformly without acceleration.

The electromagnetic field equations include a set of relations connecting the curl of the magnetic field with the total vector current density (conduction plus displacement). These relations require the divergence of the total vector current density to vanish. In m.k.s. units

$$\frac{\partial}{\partial z} \left[i'(z, t) + \epsilon_0 \frac{\partial}{\partial t} E'(z, t) \right] = 0, \quad (4)$$

or

$$z^n \left[i'(z, t) + \epsilon_0 \frac{\partial}{\partial t} E'(z, t) \right] = J(t). \quad (5)$$

Equation (5) states that the total current is independent of position along the beam.⁵ In the absence of an hf return path the total current and also the curl of the magnetic field must vanish. But the curl of the electric field is zero since the field is purely longitudinal. Consequently a magnetic field is not required to com-

⁵ F. B. Llewellyn and A. E. Bowen, Bell Sys. Tech. J. 18, 280 (1939).

plete the solution. The electromagnetic field equations possess solutions for systems *a*, *b*, and *c* containing only a longitudinal component of electric field and no magnetic field.

Continuing the analysis under the condition that no hf return path exists we get

$$\epsilon_0 E'(z, t) = - \int_{(z-z_1)/V_0}^t i'(z, t') dt', \quad (6)$$

or

$$\omega \epsilon_0 E(z, \theta) = - \int_0^\theta i(z, \theta') d\theta'. \quad (7)$$

Note that $E(z, 0) = 0$ as is required by the uniform motion of the center of the bunch. With the help of Eqs. (2) and (3) the field can be expressed in the form

$$E(z, \theta) = - \frac{i_0}{\epsilon_0 \omega} \left(\frac{z_1}{z} \right)^n X(z) \sin \tau. \quad (8)$$

The value of the electric field prescribed by Eq. (8) must now be inserted into the equation of motion:

$$z = - \frac{e i_0}{\epsilon_0 \omega m} \left(\frac{z_1}{z} \right)^n X(z) \sin \tau. \quad (9)$$

Equation (9) can be integrated formally with respect to t thus expressing z as a function of τ and t . A more convenient form is provided by the substitution $dt = dz/v_0$ which is an exact relation for electrons at the center of a bunch and a good approximation (since $\alpha \ll 1$) for all other electrons. As a result of this substitution the integral form of Eq. (9) begins to resemble

TABLE I.

System	*
Plane (a)	$\frac{1}{2} \mu a k' \sin k(z-z_1)/k$
Cylinder (b)*	$\frac{\pi}{2} \mu a k' (z z_1)^{1/2} [N_1(2(\Delta z_1)^{1/2}) J_1(2(\Delta z_1)^{1/2}) - N_1(2(\Delta z_1)^{1/2}) J_1(2(\Delta z_1)^{1/2})]$
Sphere (c)	$\mu a k' \left(\frac{z z_1}{1-4\Gamma} \right)^{1/2} \sinh \left\{ \frac{1}{2} (1-4\Gamma)^{1/2} \ln \frac{z}{z_1} \right\}$

* The Bessel functions $J_1(n)$ and $N_1(x)$ are tabulated in Jahneke-Emde, pages 137 and 191.

Eq. (2). There finally results

$$\theta = \tau - \left[\frac{1}{2} \alpha \mu k' (z-z_1) - \frac{i_0 e z_1^n}{\epsilon_0 m V_0^3} \times \int_{z_1}^z (z-z_1) \frac{X(z')}{z'^n} dz' \right] \sin \tau, \quad (10)$$

in which $k' = \omega/v_0$. The consistency of the calculation now depends on the identity of Eqs. (2) and (10). These equations become identical if the bunching parameter $X(z)$ can be determined as a solution of the integral equation

$$X(z) = \frac{1}{2} \alpha k' \mu (z-z_1) - \frac{i_0 e z_1^n}{\epsilon_0 m V_0^3} \times \int_{z_1}^z (z-z') \frac{X(z')}{z'^n} dz'. \quad (11)$$

Before integrating Eq. (11) it is convenient to introduce a bit of notation. Let

$$\begin{aligned} h^2 &= \frac{i_0 e}{\epsilon_0 m v_0^3}, & \text{System } a, & \quad n=0, \\ \Lambda &= \frac{i_0 e z_1}{\epsilon_0 m v_0^3}, & \text{System } b, & \quad n=1, \\ \Gamma &= \frac{i_0 e z_1^2}{\epsilon_0 m v_0^3}, & \text{System } c, & \quad n=2. \end{aligned} \quad (12)$$

The parameters h and Λ have the dimensions of a reciprocal length while Γ is a pure number. To obtain numerical values consider the special numerical examples:

System (a)

$$i_0 = 1 \text{ ampere/cm}^2, \quad V_0 = 2500 \text{ volts}, \\ h = 0.868/\text{cm};$$

System (b)

$$2\pi i_0 z_1 = 1 \text{ ampere/cm}, \quad V_0 = 2500 \text{ volts}, \\ \Lambda = 0.120/\text{cm};$$

System (c)

$$4\pi i_0 z_1^2 = 1 \text{ ampere}, \quad V_0 = 2500 \text{ volts}, \\ \Gamma = 0.060.$$

Numerical results for other values of current and voltage are readily found since all three

quantities defined in Eq. (12) vary directly as the current and inversely as the three-halves power of the beam voltage.

Equation (11) is best discussed by transforming it into a differential equation. Successive differentiations with respect to z yield

$$\left(\frac{d^2}{dz^2} + h^2\right)X(z) = 0, \quad \text{System } a,$$

$$\left(\frac{d^2}{dz^2} + \frac{\Lambda}{z}\right)X(z) = 0, \quad \text{System } b,$$

$$\left(\frac{d^2}{dz^2} + \frac{\Gamma}{z^2}\right)X(z) = 0, \quad \text{System } c.$$

The boundary conditions are

$$X(z_1) = 0,$$

$$\frac{dX}{dz} = \frac{1}{2}\mu\alpha k', \quad z = z_1, \quad k' = \omega/v_0. \quad (13)$$

The final results for $X(z)$ are listed in Table I.

Consider now the limiting process $z_1 \rightarrow \infty$ while $z - z_1$ and i_0 remain fixed. Under these conditions systems b and c are indistinguishable from system a . With the help of the relations $\Lambda = h^2 z_1$, $\Gamma = h^2 z_1^2$ and the asymptotic values of J_1 and N_1 (Jahnke-Emde, page 138) the second and third lines of Table I are readily reduced to the first line.

The assumption $z \geq z_1$ has been implicit in the preceding calculation. However it is necessary only to reverse the signs of i_0 and v_0 to obtain the description of systems in which the beam converges toward the axis (b) or the center (c) of symmetry. This change leaves the expressions in Table I invariant in form; they are also negative valued when $z - z_1$ is negative and not too large. It is likely that convergent beams may prove useful in the development of multipliers.

A peculiar property of the spherical beam emerges from the examination of $X(z)$ in Table I. The bunching parameter of a divergent spherical beam is a monatomic increasing function of z when the current is limited by the inequality $4\Gamma < 1$ (this requires a total current below 4.16 amperes at $V_0 = 2500$ volts).

A further point of interest is the total voltage

through which the electron falls while traversing the buncher grid space and the drift space:

$$\begin{aligned} \delta V &= \int_{z_1}^z E(z', \theta) dz' + \mu V_1 \sin \tau, \\ &= \frac{2V_0}{k'} \frac{dX}{dz} \sin \tau. \end{aligned} \quad (14)$$

Equation (14) states that the velocity modulation vanishes at points where the bunching parameter attains maximum or minimum values.

It will be observed that the restriction $|X(z)| \leq 1$ holds for all values of z if α falls below a critical value. Otherwise $X(z)$ may exceed unity for some range $z_2 < z < z_3$ and the analysis fails beyond $z = z_3$. In physical terms z_2 is the point at which overtaking or crossing of orbits begins to occur. No progress has been made in evaluating the effect of space charge beyond the initial crossover point.

A simple, but far-reaching, generalization of the preceding analysis may be obtained by the substitution of an arbitrary periodic function $f(\tau)$ for $\sin \tau$. Let

$$f^*(\tau) = \sin \omega t + \sum_n a_n \sin(n\omega t + \Phi_n) \quad (15)$$

represent the time dependence of the voltage across the buncher grids. Then

$$\begin{aligned} \mu(s)f(\tau) &= \mu(s) \sin \tau \\ &+ \sum_n a_n \mu(ns) \sin(n\tau + \Phi_n'). \end{aligned} \quad (16)$$

In particular the right-hand member of Eq. (16) may represent the sawtooth function or an approximation to the sawtooth function. If

$$a_n \mu(ns) = \frac{(-1)^{n+1}}{n} \mu(s), \quad \Phi_n' = 0 \quad (17)$$

overtaking or crossing of orbits first occurs at $X(z) = 2$ coincident with the formation of a perfect bunch (all particles passing the center of the buncher grid space during one cycle pass the fixed point defined by $X(z) = 2$ at the same instant). In this case the former restriction $|X(z)| \leq 1$ should be replaced by $|X(z)| \leq 2$.

Here and There

Harvard Computation Laboratory

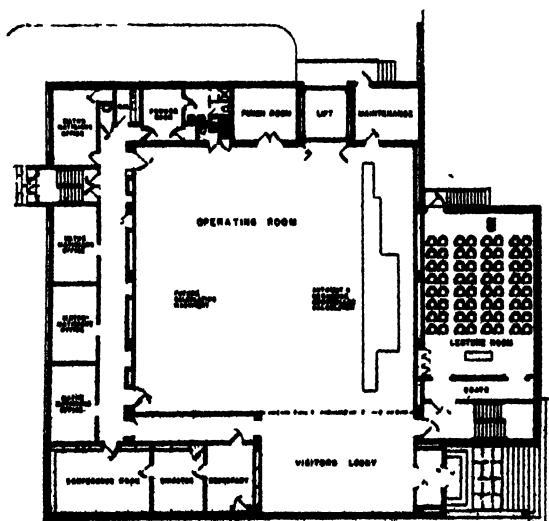
Harvard University's gigantic mechanical brain, the Automatic Calculator, will soon be moved to a new two-story building on the campus. The structure, to be known as the Computation Laboratory of Harvard University, of brick, concrete, and steel, functionally modern in design, is expected to be ready in October.

The Calculator has been doing highly secret war work since it was given to the University by the International Business Machine Corporation in August 1944, and it will continue to operate under a naval contract for two more years. By the terms of this contract, the Calculator now works twenty hours a day, every day in the week, for the Navy, and four hours a day for Harvard. It is under the direction of one of its inventors, Howard H. Aiken, associate professor of applied mathematics.

A constantly whirling device made up of gears, counters, switches, shafts, and control circuits, with 500 miles of wire, 2,000,000 wire connections, and 1464 ten-pole switches in its still-framed body, the Calculator will attain new heights of accomplishment during the next few years. Atomic physics, radio research, optics, electronics, and astronomy are only a few of its fields. With infallible accuracy and great speed, it solves problems which no scientist, however gifted, could live long enough to finish.

The new building will give added scope to these and other activities of the Calculator. Research into the design of instruments for the benefit of science will be furthered through the use of larger quarters, and instruction of graduate students under the new department of Engineering, Sciences and Applied Physics will be increased. Civilian use of the machine will include instruction for students in the design and use of large scale calculators.

In its new home the Calculator will be observed by visitors from a special gallery with a window two stories



Computation Laboratory for Harvard University, first floor plan.

high. Since it was set up at Harvard, the Calculator has been viewed by more than 2000 visitors, including scientists from Russia, China, France, Brazil, Mexico, Britain, Sweden, South Africa, Cuba, and Holland.

New Appointments

Several appointments were recently announced by the National Bureau of Standards. Dr. John G. Thompson is now Chief of the Metallurgy Division, and William F. Roeser is Assistant Chief. George A. Ellinger has been designated Chief of the Optical Metallurgy Section, and Harold E. Cleaves is new Chief of the Chemical Metallurgy Section. Dr. Wilmer Souder has been appointed Chief of the Metrology Division, formerly the Division of Weights and Measures, and there are three new section chiefs within this Division: Dr. I. C. Schoonover, Dental Materials Section; Dr. Peter Hidnert, Thermal Expansion Section; and H. Haig Russell, Large-Capacity Scales Section.

Recent Awards

Benjamin Goldberg of New York has been awarded the Exceptional Meritorious Civilian Service Award as a result of his work for the past five years as an optical physicist attached to the technical staff of the Army Engineer Board, Fort Belvoir, Virginia. He won this recognition for his contribution to the technique of metal mirror reflector manufacture and for designing special reflectors for lighting and for other optical devices, particularly for searchlights used extensively by the U. S. Army during the war.

The I.E.S. Medal, given by the Illuminating Engineering Society "in recognition of meritorious achievement which has conspicuously furthered the profession, art, or knowledge of illuminating engineering" has been awarded to Eugene C. Crittenden, Chief of the Electrical Division, National Bureau of Standards, and will be presented to him at the Society's convention in Quebec in September. The most outstanding of Mr. Crittenden's notable achievements in illuminating engineering are his contributions to the development and establishment of acceptable standards and units for the measurement of light.

Army Purchases Atom Smasher

The Manhattan District Project of the Army Engineers, which developed the atomic bomb, has ordered a 100,000,000-volt atom-smashing x-ray machine from the General Electric Company. The machine is being obtained by the Clinton Laboratory at Oak Ridge, Tennessee, which is operated by the Monsanto Chemical Company for the Manhattan District Project.

Auburn Foundation Receives Grant-in-Aid

A Frederick Gardner Cottrell Special Grant-in-Aid has been made to the Auburn Research Foundation of the Alabama Polytechnic Institute to support research on the crystal structure of sugars and their derivatives. The grant totals \$7500.

Journal of Applied Physics

Volume 17, Number 11

November, 1946

Theoretical Physics in the Metallurgical Laboratory of Chicago *

BY E. P. WIGNER

Clinton Laboratories, Oak Ridge, Tennessee

DR. DARROW has suggested that a short introduction may enhance the usefulness of the more specialized papers to be presented to the Society by members of the old Theoretical Physics group of the Metallurgical Laboratory (Plutonium Project). Because the presentation of these papers will have to be very brief and because some of our collaborators could not present abstracts in time for the meetings, I was most happy to follow our Secretary's invitation to summarize our work in general terms.

There were, during the period extending from about the middle of 1942 until about the middle of 1945, that is for about three years, approximately twenty theoretical physicists assembled on the fourth floor of Eckart Hall who worked on those problems of chain reacting units which could be solved or at least attacked with the methods of theoretical physics. The membership of the group varied somewhat but I am happy to say, and do say it with a great deal of pride, that we formed a happy family and that I hardly remember a disagreement on non-technical points.

Most of our work was on very urgent problems and only a small fraction of it will bear publication in scientific periodicals. Our group had to do a great deal of engineering work, so much so that the calculation of liquid flow rates became to be considered to be part of our responsibilities. We

calculated more than one I beam dimension, among similar matters, although we felt somewhat out of place when we did that. In addition, we had to maintain numerous contacts with the experimental groups and to take a lively interest in matters ranging from the fabrication and corrosion of aluminum tubes to the radioactivity induced in oxygen by neutron absorption. All this work was necessary and I do not hesitate to say in retrospect that our policy of assuming some of the functions which are usually reserved for engineers proved most useful. In the first place the Plutonium Project was not well provided with engineers in its early days—it had only two or three design engineers at the time when our plan for the *W* unit (which was later erected at Hanford) was virtually ready**; in the second place it was important, as in any new field such as that of chain reacting units was in 1942, that there be a few people at least who are sufficiently familiar with the whole picture to know about every difficulty and how it may be overcome. This second point will be very important in connection with Dr. Daniels' pile which is to be erected at Clinton. I fear that the present arrangement, which gives those who are

** It may be interesting to remark that the date on the *W* report is January 9, 1943 and that it was issued just 42 days after the first nuclear chain reaction was experimentally established by E. Fermi and his collaborators (December 2, 1942). Of course, the relevant dimensions and sizes, etc., of the *W* pile had to be fixed much prior to that date, although, of course, they had to be verified experimentally later.

* An address presented to the American Physical Society at the Chicago Meeting, June 22, 1946.

mainly interested in this pile even less responsibility than we had with respect to the *W* pile, will not help in meeting the time schedule which we read about in the newspapers. That time schedule was quite optimistic in the first place and I do not believe that the pile will be in operation before 1948.

As I said, knowing all details of the *W* plan was our main function and it took up most of our time. Practically all senior members of the group participated in it in one form or another although most of the burden was carried by Messrs. Friedman, Ohlinger, Weinberg, and Young and Miss Way. In addition, in the second year, Mr. J. A. Wheeler was transferred to Wilmington to give direct help to the DuPont Company. While work on the *W* plans was our most important function in the first two years, at the same time it was the one which is least suitable for being reported in public. In later years, work on the power production occupied most of our interest. However, this subject, to which Mr. G. Young has contributed so decisively, is still restricted from public discussion so that my report will deal, essentially, only with incidental studies which we have made. I shall deal with these under four headings:

1. Elementary theory of nuclear chain reactions.
2. More detailed theory of chain reactions.
3. Effect of radiation on matter.
4. Studies in theoretical physics

I shall give an outline of our work in these fields in the above order.

1. ELEMENTARY THEORY OF NUCLEAR CHAIN REACTIONS

The great surprise about nuclear chain reactions was the ease with which they could be established. Szilard's paper of January 1940 already describes a workable arrangement. Our own early work in this field was not based on Szilard's paper but on Fermi's work, the concepts of which are less intricate than Szilard's. Ideas similar to Fermi's were developed also by others, notably by v. Halban; moreover, the whole work was duplicated, apparently without any major deviation, also by the German nuclear physicists. In a chain reaction of the kind considered by us, uranium nuclei undergo fission and liberate

neutrons. These neutrons are first fast but are soon slowed down by the moderator which is carbon (graphite) in our case. After being slowed down, the neutrons still diffuse around for a period of time before being absorbed. Most of them are absorbed by uranium, which then undergoes fission and emits the neutrons of the next generation. The ratio of the number of neutrons in one generation to the number of neutrons of the preceding generation is called multiplication constant and was usually denoted by k . Fermi's theory divides the problem of multiplication constants and critical sizes into two parts. The first problem is the calculation of the multiplication constant in an infinite medium, k_{∞} , which is usually referred to briefly as the multiplication constant. It depends only on the geometry and the materials of the chain reacting system and gives the ratio of the numbers of neutrons in successive neutron generations under the assumption that the same materials, arranged in the same geometry, extend all over infinite space.

The second problem is the calculation of a critical length which does not depend on the inner structure of the chain reacting unit but only on its size and shape. This critical length, or its reciprocal κ , permits one to calculate, from k_{∞} , the second kind of multiplication constant, k_{eff} . This gives the ratio of the numbers of neutrons in successive generations in a finite pile. This second kind of multiplication constant is, of course, the relevant one from a practical point of view. It depends not only on the materials and their arrangement into a lattice which determine k_{∞} , but depends also on the actual extension of the lattice, i.e., the size and shape of the pile. In a steadily running pile k_{eff} is always 1 and it exceeds 1 only when the power of the pile is increased, e.g., during startup, and then only very little. The k_{eff} is always smaller than k_{∞} because in an actual, finite pile some of the neutrons of every generation diffuse out of the pile and do not contribute to the next generation. No such "leakage" exists in an infinite pile.

It would seem that only the effective multiplication constant has real significance but it turns out that the calculation of k_{∞} is an almost necessary preliminary for the calculation of k_{eff} .

I will only sketch the calculation of k_{∞} which

is already given, in principle, in the Smyth report. In order to calculate the number of neutrons of the next generation produced by one neutron of the present generation, one may start at the birth of one neutron. This occurs in the uranium lumps and the neutron has, originally, considerable velocity. As a result, it will be able to induce fission not only in the U^{235} nuclei but, what is more important because of their larger numbers, also in the U^{238} atoms. Competing with this process are the process of inelastic scattering by U atoms by which the original neutrons may be slowed to a velocity below the fission threshold¹ of U^{238} , and the process of escape of the neutrons from the U lump into the moderator.

The importance of fast fission was recognized by Szilard and his collaborators. The rest of the factors making up k_{∞} were all recognized before, and are contained also in Fermi's considerations.

Let us assume that the original neutron generates $\epsilon=1$ further neutrons by fast fission. As Fermi has mentioned in his address to this meeting $\epsilon-1 \approx .03$ so that we arrive with $\epsilon \approx 1.03$ neutrons just below the fission threshold. Most of these neutrons diffuse out into the moderator and are slowed down to thermal energies. Some of them occasionally enter the uranium and are absorbed there by one of the numerous resonance levels of the U^{238} . These absorptions do not lead to fission and constitute an actual loss of neutrons. The importance of this process has been already recognized by N. Bohr in 1940, and others. Only when the neutron has lost sufficient energy to be below the lowest resonance level of U^{238} —which is, according to data in the literature² at about 5 ev—is it safe from this fate. The probability that a neutron will escape resonance capture is usually denoted by p . It is a number smaller than 1. As a result of the resonance absorption, we arrive with ϵp neutrons below the energy of 5 ev instead of the ϵ neutrons which we had just below the fast fission threshold.

It would lead too far to describe the actual calculation of p . Among all the processes which contribute to the chain reaction, the resonance absorption is the only one which was not really

understood when we started our work. S. M. Dancoff and I were the ones who were most interested in the physical principles which determine the resonance absorption of macroscopic bodies, but ideas similar to ours were developed also by others. The actual calculation of p was described by R. F. Christy, A. M. Weinberg, and myself, although many others, including H. L. Anderson, contributed to it. The material constants necessary for the calculation were measured by Creutz, Jupnik, Snyder, and R. R. Wilson in Princeton, and later by Mitchell's group at the University of Indiana.

We now have ϵp neutrons with an energy below the resonance levels of uranium. According to theory, they will be slowed down to thermal energies by the moderator. After that, they will be absorbed, some of them by the moderator and the impurities present in the pile, some of them by the uranium. Fermi denotes this last fraction by f so that, altogether, ϵpf thermal neutrons are absorbed by the uranium, giving

$$k_{\infty} = \epsilon pf \eta \quad (1)$$

secondary neutrons, η being the number of fission (fast) neutrons produced in the uranium per thermal neutron absorbed. The principles for the calculation of the "thermal utilization" f were established independently by Fermi, Placzek, and our group. The formulae which we used were derived by Christy, Mrs. Monk, Plass, and myself in a way which is similar to the calculation of wave functions in metals by the cellular method.³

On the whole, the calculation of the multiplication constant for an infinite lattice is quite "straightforward" and one of the great surprises of the Plutonium Project was how easy it was. Mr. G. N. Plass and myself happened to be the ones who attempted to calculate the "optimal lattice" (i.e., the lattice with the highest k_{∞}) early in 1942. Although the physical constants were not known at that time too accurately, the dimensions we obtained (later incorporated in the first chain reacting unit) are now believed to give a k_{∞} just $\frac{1}{2}$ percent short of the k_{∞} of the real optimal lattice. We are quite convinced that

¹ Haxby, Shoupp, Stephens, and Wells, *Phys. Rev.* **57**, 1088A (1940); **58**, 199A (1940).

² H. L. Anderson, *Phys. Rev.* **57**, 566 (1940).

³ Cf. e.g., F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), Chap. IX.

any reasonably competent people would have arrived at the same results. In later years, computations of k_{∞} were much facilitated by diagrams prepared for this purpose by Mrs. Monk and Mrs. Uchiyamada, under Professor Wheeler's direction. The calculation of k_{∞} was extended to all sorts of lattices, containing heavy and ordinary water, etc. Most of this work was done by A. M. Weinberg and his collaborators, Mrs. Monk, Mr. Plass, Mrs. Uchiyamada, Mr. Stephenson and others. Qualitatively, the results were quite similar for all systems considered.

In spite of this, the properties which make a lattice optimal are not very simple. One may note that it is good if the high energy neutrons remain in the uranium to give a high ϵ . On the other hand, it is best if the lower energy (resonance) neutrons keep out of the uranium as much as possible so that p may remain reasonably close to 1. Again, the thermal neutrons should return to the uranium to give a high f —as close to 1 as possible. These conflicting requirements determine the geometry of the optimal lattice, i.e., give the ratio of the amounts of moderator and uranium as well as the lattice constant. However, even relatively large deviations from the optimal dimensions do not decrease k_{∞} to a very great extent.

The foregoing describes the calculation of k_{∞} . Although Fermi has given a method for calculating k_{eff} from k_{∞} , I will not give his method here but will turn to the more advanced theory which permits a direct calculation of k_{eff} .

2. MORE DETAILED THEORY OF CHAIN REACTIONS

The more detailed theory of chain reactions should provide more accurate methods both for the calculation of k_{∞} and also for the calculation of k_{eff} . However, as far as k_{∞} is concerned, only few improvements were made. None of these improvements occurred in the calculation of ϵ and p , only one occurred in the calculation of f .

The behavior of "thermal" neutrons in a moderator-uranium lattice is far from simple. Evidently, it would take infinitely many collisions to establish real thermal equilibrium between the neutrons and the moderator, and in a well-designed lattice the neutrons will be absorbed by the uranium after a relatively small

number of collisions. As a result, the energy spectrum of the neutrons will remain quite complex and their average energy will stay considerably above $\frac{1}{2}kT$. This average energy will be different even at different points of the lattice. The actual energy distribution will be influenced by the absorbing power of the material as well as by its moderating power. The latter is influenced in turn by the atomic weight of the moderator, by Fermi's chemical binding effect and by the crystalline nature of the moderator which gives a considerable anisotropy to the scattered (refracted) neutrons.

The only serious attempt to take these factors (excepting the crystal effect) into account is due to E. Teller and his collaborators, mainly N. Metropolis and P. Morrison. A more rigorous but much more formal attempt later by E. J. Wilkins and myself did not contribute much to the qualitative picture. Wick reported on Thursday about some work which he did on this problem. Teller's work gave, at least, an approximate measure for the difference in the effective temperature of the neutrons and the moderator. In spite of this we are far from having an adequate knowledge of the energy spectrum of the neutrons in a chain reacting unit.

Moreover, the problem of calculating f remains far from being simple even if the energy spectrum of the neutrons is known. It is, in fact, quite complicated even if one assumes that all the neutrons have the same energy. The reason for this is that the ordinary diffusion theory proves to be quite inadequate. G. Placzek carried out the most accurate calculations for the diffusion of monoenergetic neutrons. Some of his results were obtained also by German and Italian theorists⁴ and published. Our work along this line was not pushed with much vigor because we were, perhaps, too well aware of the inadequacy of the model which uses monoenergetic neutrons. Actually, there is evidence that the errors in our primitive diffusion equations are quite substantial and in the direction indicated by Placzek's work.

There is no relevant difference between the fast effect in a finite and infinite lattice. However, the probability p that a neutron with an energy

⁴G. C. Wick, *Zeits. f. Physik* 121, 702 (1943).

just below fission threshold should become a thermal neutron is smaller in a finite lattice than in an infinite one because in addition to being captured by the uranium, some neutrons will be lost from a finite lattice by "leaking" (diffusing) out of it. This leakage was calculated by Fermi and his co-workers⁵ even before fission was discovered. For a finite lattice, their work gives

$$p_{eff} = p \exp(-\tau\kappa^2), \quad (2)$$

where τ is one-sixth the mean square distance, in an infinite lattice, between the point where the neutron originated and the point where it becomes thermal. The quantity κ^2 is the ratio $-\Delta n/n$ where n is the average of the neutron density over a lattice cell and will be discussed presently.

According to (2) the effective p is smaller than it would be for the constant n of an infinite lattice, i.e., for $\kappa=0$. The leakage depends on the "age" τ , which on its turn increases with increasing mean free path of the neutrons in the moderator and with the number of collisions which are necessary to slow down the neutrons to thermal energies. The quantity τ , and hence the leakage, is smallest in a water moderated pile and much greater in a graphite moderated pile.

Just as the fraction of neutrons which are slowed down to thermal energies in the pile is, because of the leakage, smaller in a finite than in an infinite pile, so is the fraction of thermal neutrons absorbed by the uranium decreased by the escape of some of the thermal neutrons from the pile. The equation analogous to (2) is

$$f_{eff} = f(1 + L_p^2 \kappa^2)^{-1}. \quad (3)$$

The significance of κ in (3) is the same as in (2), that of L_p^2 similar to that of τ in (2): L_p^2 is one-sixth of the mean square distance in an infinite lattice between the point where the neutron becomes thermal to the point to which it has diffused when it gets absorbed. L_p is also called the diffusion length of thermal neutrons in the lattice because the n decreases with an exponential relaxation distance L_p

$$n \sim \exp(-x/L_p) \quad (4)$$

in a region in which no thermal neutrons are

⁵ E. Fermi and F. Rasetti, *Ricerca Scient.* 9, 472 (1938); G. Placzek, *Phys. Rev.* 69, 423 (1946).

produced. G. N. Plass showed, by means of a calculation which is similar to Bardeen's work⁶ on metallic wave functions, that

$$L_p^2 = L_m^2(1-f) \quad (5)$$

is a very good approximation for L_p if L_m is the diffusion length in the pure moderator, without uranium lumps.

The condition that a lattice can maintain a chain reaction at a steady rate is that $k_{eff}=1$, i.e., that

$$k_{eff} = \epsilon p_{eff} f_{eff} \eta = 1. \quad (6)$$

Using the expressions (2) and (3), this becomes

$$\epsilon p f \eta \exp(-\tau\kappa^2)(1 + L_p^2 \kappa^2)^{-1} = 1$$

or, by (1)

$$k_{\infty} = (1 + L_p^2 \kappa^2) \exp(\tau\kappa^2), \quad (7)$$

an equation essentially identical to one already obtained by Fermi.

This last equation can be considered to be an equation for κ which, in its turn, will be seen to depend only on the size and shape of the pile. Hence (7) gives us the size of a pile if its shape and internal structure, in particular its infinite multiplication constant k_{∞} , are given.

The connection between the quantity κ and the size and shape of the pile is established by the classical equation

$$\Delta n + \kappa^2 n = 0 \quad (8)$$

in which the average neutron density n is subject to the boundary condition that it vanish at the outer boundaries of the pile. It is well known that (8) allows a solution only for definite, discreet values of κ^2 which depend on the size and shape of the region on the boundary of which n has to vanish, i.e., on the size and shape of the pile. Only for the smallest of these κ^2 is n positive throughout and this smallest κ^2 is the one which occurs in (7). Equation (8), so to say, gives an effective dimension κ^{-1} to every size and shape and (6) shows how this effective dimension affects the effective multiplication constant. If the κ of the pile, as defined by (8), is larger than the solution of (7), the pile is under critical, its effective multiplication constant smaller than 1. If the solution of (7) is larger than the κ satisfying (8), the pile is above critical.

⁶ J. Bardeen, *Phys. Rev.* 49, 653 (1936). Cf. also reference 3.

The quantity n to which (8) applies is the average neutron density, the average to be taken over a cell. Evidently, an equation, applying to such an average as (8) does, can be accurate only if this average does not change too rapidly from cell to cell. The relation of the n of (8) to the actual neutron density is similar to the relation of the macroscopic density of bodies to their rapidly fluctuating density as given by their atomistic structure. The theory of Eq. (8) is therefore called the macroscopic pile theory while the quantities of Eqs. (1) to (7) are concepts of the microscopic pile theory. Actually, (8) is only the simplest equation of macroscopic pile theory, which applies if the spatial variation of the neutron density is independent of energy. This is an important particular case but does not hold in general. For instance, most control rods absorb only low energy, thermal neutrons. The surface of a control rod is, therefore, a boundary where the density of thermal neutrons vanishes. However, the density of fast neutrons does not vanish at the surface of the control rod and the densities of fast and of slow neutrons are not proportional any more. Problems of this nature call for more complicated equations than (8). The most important results toward the solutions of these problems are due to Messrs. F. L. Friedman, A. M. Weinberg, and J. A. Wheeler.

Even the simple Eq. (8) raises a number of interesting problems. If the shape of the pile is at all complicated—which is almost invariably the case if the chain reacting material is liquid—the solution of (8) could be obtained only by perturbation methods. Some of these show a remarkable similarity to the Rayleigh-Schrodinger method with which we are familiar from its application to quantum-mechanical problems. We owe many interesting results on (8) to Messrs. F. Murray, L. W. Nordheim, and H. Soodak.

A good part of the work in this connection is too special to be taken up in detail. Another part will be dealt with by the speakers following my address so that I may close the subject of the calculation of multiplication constants and critical sizes. I would not like to do this before emphasizing that, in my opinion, a good deal of work remains to be done in this field. In particular, the behavior of "thermal" neutrons in the

pile and the transition from fast to thermal energies requires further clarification both from the experimental and from the theoretical side. But there remain interesting details to be worked out in almost any part of the theory. There are, also, some problems which have already commanded considerable attention but which I have not even touched. Chief among these is the change in the neutron densities with time if Eqs. (6), (7) are not exactly fulfilled and the pile is either below or above critical. Messrs. R. F. Christy, L. W. Nordheim, and J. E. Wilkins were particularly active in this field.

3. EFFECT OF RADIATION ON MATTER

The radiation densities, both γ and neutron, are higher in a plutonium producing pile than can be maintained outside the pile for extended periods. The effect of these radiations on the structure of materials was one of our early concerns from the theoretical point of view. The experimental work was carried out in the Chemistry Division. Dr. M. Burton reported at the Atlantic City meeting of the American Chemical Society about his, his collaborators' and Dr. J. Franck's work on the subject. On the theoretical side, M. Goldberger, R. S. Mulliken, and F. Seitz shared my interest in the subject which still has some aspects about which we cannot talk freely.

Clearly, the collision of neutrons with the atoms of any substance placed into the pile will cause displacements of these atoms. If the substance is a chemical compound, the displacement will result in chemical changes which were, of course, investigated already before chain reacting units came into being and are summarized, e.g., in the booklet of Lind.⁷ All these changes are much more intense in the pile owing to the more intense radiation. But substantial effects can be expected in elementary substances also. The matter has great scientific interest because pile irradiation should permit the artificial formation of displacements in definite numbers and a study of the effect of these on thermal and electrical conductivity, tensile strength, ductility, etc. as demanded by theory. One may expect that

⁷ Cf. e.g. S. C. Lind, *Chemical Effects of α Particles and Electrons* (Chemical Catalogue Company, New York, 1928).

studies of the solid state, particularly of the structure sensitive properties, will be greatly stimulated by the additional experimental facility given by the pile.

Before a final interpretation of the experimental results can be made, our knowledge of the ranges of low energy ions will have to be extended. It is on this subject that most of Messrs. Goldberger's and Seitz's work was concentrated and you will hear from them presently. A good deal of the rest of our work was speculation which will either be confirmed or refuted by future experiments.

4. THEORETICAL PHYSICS

As I emphasized before, real theoretical physics was always somewhat of a stepchild. This situation hardly could be remedied in view of our many pressing obligations. We tried to free Mr. S. M. Dancoff as much as possible from the pressure of urgent work and he carried out several investigations which are of considerable general interest. There is only one other function which we did not permit to be pushed into the background and this was the problem of keeping tables of nuclear constants up to date. Mrs. Uchiyamada was in charge of these tables but she received a good deal of help from the Project Information Department, in particular from Mr. Goldsmith who made a really brilliant contribution. We were happy to learn that these tables may be at least partially published soon.

The theoretical work of the group fell into two categories; help with the evaluation and planning of experimental work, and real theoretical work. Into the first category falls the work of Messrs. Cahn, Schweinler, Weinberg, and others on the so-called pile oscillator. This is an instrument which permits an absorber of known or unknown neutron absorbing characteristics to be put into periodic motion in the pile. The oscillation of the neutron absorber causes intensity waves to spread all over the pile. These waves are similar to the temperature waves in the earth, generated by the daily and yearly heat fluctuations of the heat input on the earth's surface. The amplitude and wave-length of the waves permits one to evaluate the characteristics

of the neutron absorbing oscillators and of properties of the pile.

The work on neutron diffraction received considerable attention on the part of Goldberger and Seitz. They interpreted and extended Weinstock's results⁸ considerably and took into account phenomena not previously considered. Their work is being continued by Mr. M. Moshinsky in Princeton. There is little need for my going into details on this question since Seitz and Goldberger will tell you about them presently.

As a last example, I would like to mention Dancoff's work on short range α -particles. This work actually started because of some acute problem which was practically forgotten by the time Dancoff took over. He noticed that the intensity of short range α 's is often anomalous in the light of Gamow's theory⁹ which stipulates that the α -particle is emitted by an excited residual nucleus. Dancoff investigated several other mechanisms among which the excitation of the residual nucleus by the α -particle after it has already penetrated the potential barrier seems to be the most important. These theoretical investigations have now received added interest in view of Chang's¹⁰ experimental results. Chang has discovered Dancoff's mechanism independently. Dancoff will tell you about this phase of his work in the afternoon.

I would not like to close my review of the work of our group without expressing my sincere thanks to all members of the group for their most unselfish and loyal cooperation. I have to extend my apologies to those whose work I may have slighted. As I said before, we had a very concrete objective in mind during the course of the work and the most important problems solved were not always the ones which now appear most worth remembering. We have, however, encountered a good many interesting problems, several of which will bear a great deal of further study. A glimpse at these problems, together with the knowledge of the importance of our aims, contributed a great deal to making the relations between our group cordial.

⁸ R. Weinstock, *Phys. Rev.* 65, 1 (1944).

⁹ Cf. G. Gamow, *Structure of Atomic Nuclei* (Clarendon Press, Oxford 1937), p. 104 ff.

¹⁰ W. Y. Chang, *Phys. Rev.* 69, 60 (1946).

Contributed Original Research

Composition and Diffraction Effects in X-Ray Microradiographs

R. SMOLUCHOWSKI,* C. M. LUCHT, AND J. M. HURD
General Electric Company, Schenectady, New York

(Received June 1, 1946)

MICRORADIOGRAPHY is not a new art. However, its systematic application to various problems is rather recent. The principal idea of the method is extremely simple: A thin sheet of a material is placed on a photographic plate and exposed to x-rays. The resulting image on the photographic plate is enlarged or examined under a microscope; the art consists in obtaining reproducible results, which can be unambiguously interpreted. The early work in the field of microradiography was primarily in the field of biology¹ and geology.² The first to use it for the study of metals were Fournier³ and later Trillat.⁴ The latter author especially has obtained interesting examples of the distribution of various elements in alloys as well as of internal defects. Since then a number of authors have improved and described various aspects of this method.⁵ The purpose of the present investigation⁶ was not so much to improve the technique as to make a careful comparison with results obtainable with a conventional light microscope and to look for and interpret new *diffraction* effects.

THEORETICAL

The theory of concentration effects is very simple and has been discussed in numerous places.⁵ Only the fundamental facts will be recalled here: Every chemical element has its characteristic x-ray absorption spectrum. Thus by a proper choice of x-ray wave-length (anode

material and voltage) one can produce a difference in absorption of various constituents of an alloy. Concentration gradients in the direction perpendicular to the x-ray beam show up as gradients in blackening. A uniform thickness of the metallic foil, smaller than the size of the concentration non-uniformities, is naturally assumed. Apart from a careful consideration of the wave-lengths and absorption coefficients, two facts should be borne in mind in the interpretation of the results; First, that a microradiograph is the projection of the whole thickness of the foil on the photographic plate and not only of its surface and, secondly, that except in the proximity of an x-ray absorption edge, the absorption does not depend upon the chemical state of the atom.

The theory of diffraction effects is equally simple. However, it was not described previously in this connection and so a short description may not be out of place. Let us assume that the shaded elongated areas in Fig. 1 indicate cross sections through various grains in a metallic foil placed perpendicularly to the plane of the drawing, and that the lines themselves indicate orientation of a set of crystallographic planes in each grain. Monochromatic x-rays passing through the foil in the direction indicated by the arrows will undergo absorption according to the usual

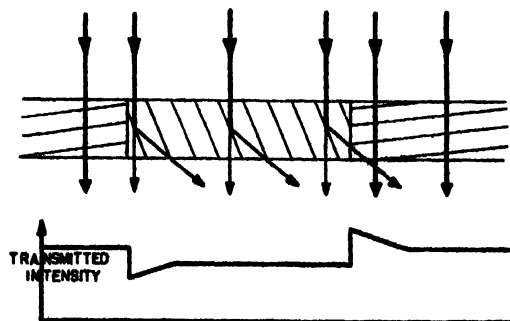


FIG. 1. Diffraction effects on grain boundaries.

* Now at the Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

¹ A. Dauvillier, *Comptes rendus* 190, 1287 (1930).

² P. Goby, *Comptes rendus* 156, 686 (1913).

³ F. Fournier, *Rev. Metallurgie* 35, 349 (1938).

⁴ J. J. Trillat, *Rev. Scientifique* 78, 212 (1940).

⁵ For instance: G. L. Clark and W. M. Shafer, *Trans. A.S.M.* 29, 732 (1941). S. T. Gross and G. L. Clark, *Ind. Radiography* 1, 21 (1942). S. E. Maddigan, *J. App. Phys.* 15, 43 (1944). S. E. Maddigan and B. R. Zimmerman, *Trans. A.I.M.M.E.* 156, 33 (1944).

⁶ R. Smoluchowski, C. M. Lucht, and J. M. Hurd, *Phys. Rev.* 68, 100 (1945).

law relating loss in intensity to the thickness of the traversed material. However, if the direction of these x-rays happens to form a Bragg angle α with a set of planes in a particular grain, then the radiation will be diffracted and it will come out from the foil at an angle 2α with the original direction. Naturally the path of these x-rays in the foil is longer than if they were not diffracted, and so the intensity transmitted through that grain will be correspondingly weaker. Because of the exponential character of absorption this difference may be quite large. This "effective thickness" may vary from grain to grain and so various degrees of transmission can be expected.

Grains adjoining that grain which happens to diffract the x-ray beam do not need to be themselves in an orientation suitable for diffraction. In general we would expect thus two effects to be observed at the grain boundary: If in one grain the x-ray beam is diffracted away from the boundary (in Fig. 1 the left boundary), then there should be a white line along the boundary. If, on the other hand, the radiation is diffracted towards the boundary, then in the adjoining non-diffracting grain we should observe a dark line along the boundary (right boundary in Fig. 1). If two adjoining grains diffract, then, depending upon the relative orientation of the grains, their boundary may have a light line on one side and a dark line on the other.

If the photographic plate is not in direct contact with the metal foil but at some distance, say 2 mm, from it, then one observes two areas for each diffracting grain: One is a white area where the x-ray beam would have reached the plate if it were not diffracted, the other is a black area where the diffracted beam reaches the plate. If the angle of diffraction was large or the absorption high, then the dark area may not be observable because not much intensity would emerge from the diffracting grain. The white and the dark area are essentially complimentary and can be readily correlated. Knowing the distance between the sample and the plate and the distance between the two areas the angle of diffraction and thus the index of the diffracting plane can be calculated. The white and dark lines along grain boundaries are limiting cases when the two areas, the white and the dark one, almost coincide on the photographic plate.

The condition for occurrence of a Bragg reflection is quite sharp and so even a small change of orientation of the metal foil with respect to the x-ray beam should produce a change in appearance of the microradiograph and in the relative intensity of the various grains. Since intensity of diffraction depends upon the index of the crystallographic plane which is oriented at the proper angle, it follows that various gradations of intensity will be present in a microradiograph of a non-oriented polycrystalline material. An oriented material will show a simultaneous "flashing" of many grains and so microradiography provides a new tool for a study of preferred orientation. If, instead of monochromatic (in practice, filtered) radiation, the full radiation of an x-ray tube is used, then the contrasts will naturally be lower.

In the discussion here it was assumed that the crystal lattice of the metallic foil is free from strain. Naturally this does not need to be true and the changes in microradiographs produced by strains turn out to be of great interest. Their study will be the subject of a later paper.

The two effects, composition gradients and diffraction, can be distinguished in a simple manner. It is sufficient to repeat the exposure with a slightly tilted sample. If the pattern remains the same, it is caused by concentration gradients, if it changes it is a diffraction effect. Usually a change of the angle of the order of a fraction of a degree is sufficient to distinguish between the two effects. Naturally a diffraction effect may be influenced by a concentration gradient. In that case by a proper choice of anode material one can reverse the concentration effect or eliminate the diffraction effect retaining the gradient of blackening due to concentration non-uniformities. Such a procedure may be necessary, for instance, in the case of a solid solution in which a steep concentration gradient produces a gradient of the lattice constant, and thus a gradient in the condition for diffraction.

EXPERIMENTAL TECHNIQUE

The main experimental difficulty in microradiography lies in preparing the metallic foils. However, with certain practice the necessary time is about the same as for preparing regular metallographic samples. A thin slice of metal

was cut off by a cut-off wheel and then ground and polished on both sides. During these operations, the thin foils were cemented to metal blocks so as to assure a perfectly flat surface. In certain cases, when it was advisable to avoid straining the sample during polishing operations, the electrolytic polishing technique was used. The sample was again cemented to a copper block taking care to establish electrical contact between the block and the sample.

The samples were usually about .05 mm (.002") thick and were cemented over an appropriate hole in the center of a small brass disk. Figure 2 shows the construction of the camera in which the ring *R* with the sample *S* was placed in a centering depression and covered with a photographic plate. A cover *C* was fastened by means of two wing nuts.

The camera was placed within 12 cm from the focal spot of a General Electric X-Ray tube, Type CA5. The tube was mounted in such a manner that the x-ray beam emerged upwards

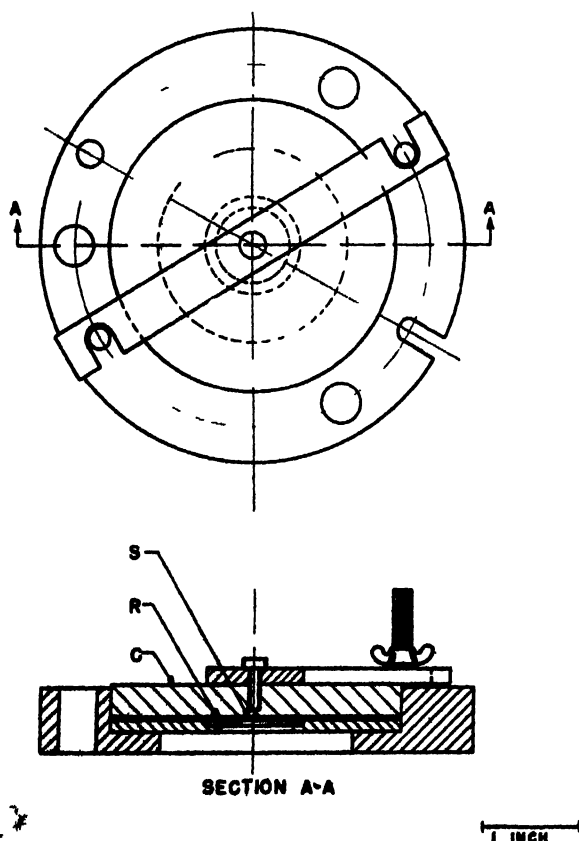


FIG. 2. Microradiographic camera. *S*—sample, *R*—supporting ring, *C*—cover.

in a vertical direction. If it was necessary to tilt the camera, then it was placed in a special support. By means of a fine dial indicator the angle of tilt could be measured to better than $\frac{1}{4}$ a minute of arc. The photographic material was standard 548-O and 649-G H Eastman Spectroscopic Plates cut to appropriate size. Both types of plates were developed for $1\frac{1}{2}$ minutes in D8 which we found to give smaller grain than the usually recommended D19. The photographic plates were examined with an ordinary microscope. If it was necessary for comparison to obtain a regular metallographic photograph of the foil, then it was etched in the usual manner. Inasmuch as etching produces surface irregu-

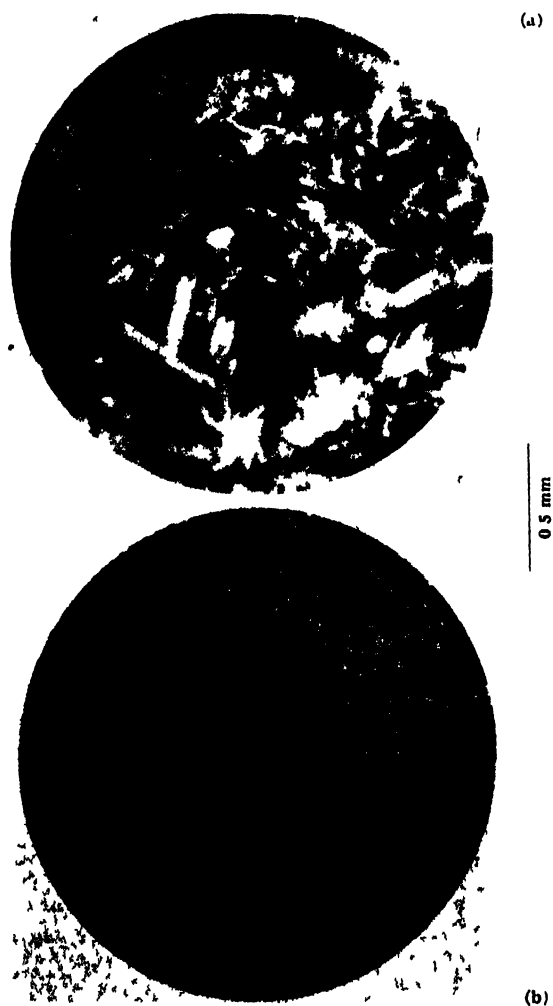


FIG. 3. Photo and x-ray microradiographs of a bearing alloy. The same area, 1.8 mm diameter. (a) x-ray, Cr anode, 20 kv. (b) photo.

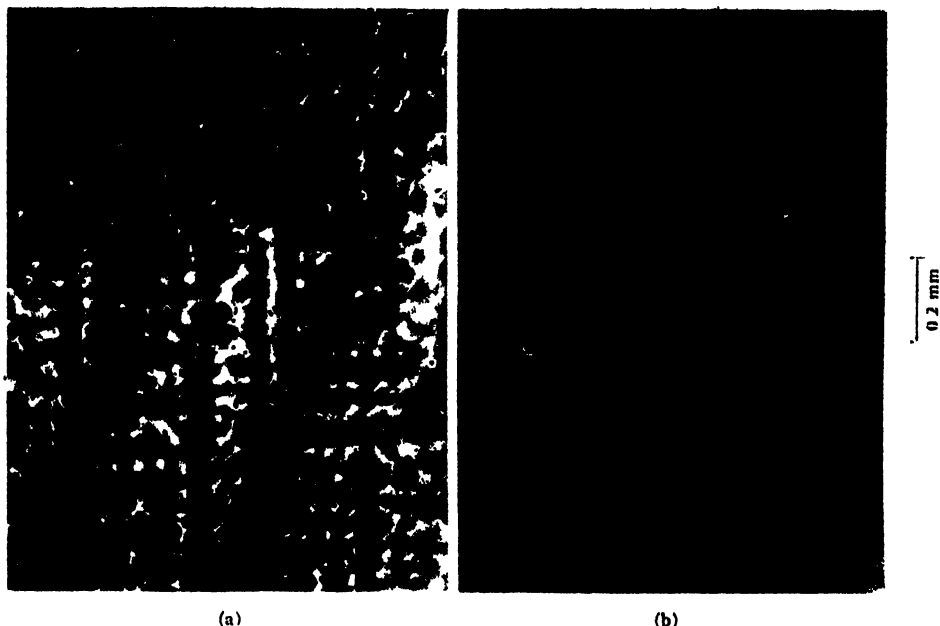


FIG. 4. Microradiographs of a cobalt alloy. The same area, 1.1×0.8 mm.
(a) Cr anode, filtered, 20 kv. (b) Mo anode, filtered, 20 kv.

larities and thus locally changes thickness of the foil, it is detrimental to a microradiographic study of other effects. For that reason all x-ray pictures were taken before the sample was etched because it was often impossible to remove the etching without destroying the thin sample. All illustrations shown in this paper are positives.

EXPERIMENTAL RESULTS

If the microradiographic technique would not show anything more than is readily visible under a light microscope, it would not have much significance. It is important, therefore, to stress the differences between the two methods.

A. Concentration Effects

1. Identification of phases

If there are several phases in a complicated alloy it is often very difficult to estimate the distribution of the various constituents among the phases. Figure 3 shows a photomicrograph and an x-ray microradiograph of a babbitt alloy (82 percent Sn, 9 percent Sb, 9 percent Cu, and some minor impurities). The *area* is exactly the *same* in both pictures. The large square precipitates are much more transparent than the rest and so they are readily identifiable as containing primarily copper. Several small precipitates do

not occur in the photomicrograph because they do not reach the surface. The light square in the lower right of the photomicrograph appears only very weakly in the microradiograph indicating that the precipitate is very shallow.

Another example of phase identification is provided in Fig. 4 which shows microradiographs of the same area of a cobalt alloy (64 percent Co, 26 percent Cr, 5 percent Mo, 2 percent Ni, 1 percent Fe, and impurities) made with filtered Cr and Mo radiations. Apart from the nicely visible dendritic structure, the finely divided phase is easily identifiable as a molybdenum rich phase. It is highly transparent to Mo radiation and very absorbing to Cr radiation. None of the other constituents would show a similar reversal. This can be seen from a comparison of the absorption coefficients per unit thickness for chromium and for molybdenum radiations. The respective values are: Co 1120, 374; Fe 910, 300; Mo 4500, 204; Ni 1290, 418; Cr 640, 213. The absorption coefficients of cobalt for both radiations are greater than the corresponding coefficients for iron and chromium and smaller than those for nickel. Only for molybdenum the first absorption coefficient is larger, the second smaller, than the corresponding values for cobalt.

In that connection it may be pointed out that

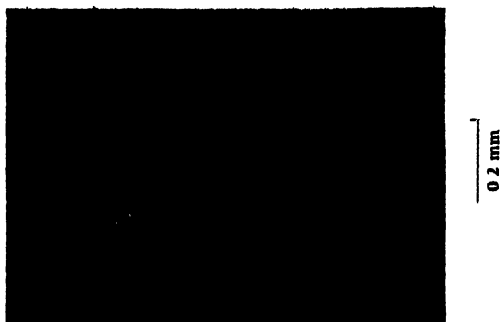


FIG. 5. Concentration gradients in a 50 percent Au, 50 percent Ag alloy. Mo anode, 35 kv. Area: 0.9×0.6 mm.

criteria for good contrast obtained on the basis of studies with unfiltered radiation are not well applicable if filtered radiation is used. Nevertheless, the following condition, suggested by Maddigan⁵ is satisfied in the microradiographs here reproduced:

$$\exp |(\mu_1 - \mu_2)t| < 0.9,$$

where μ_1 and μ_2 are absorption coefficients of the two neighboring areas and t the thickness of the layer.

2. Concentration gradients

If a metallic surface would etch uniformly, then a microscopic examination would not show any structure. In other words, what we see in such cases are only differences in etching. These differences are usually big enough to bring out boundaries between various phases in an alloy. However, if there is only a gradient of concentration, then it will show up only if there is a change in etching characteristics with concentration. This change, if any, occurs usually for a given critical concentration within the solid solution and so, instead of a gradient, there appears under the microscope an outline of the boundary between the regions of higher and of lower concentration than critical. Microradiographs are naturally very suitable for the study of concentration gradients. Figure 5 is a microradiograph of a 50:50 gold-silver alloy with very pronounced concentration non-uniformities which are due to rapid cooling during the process of crystallization.

Microradiographs also show concentration gradients surrounding separated phases. In Fig.

4 each molybdenum-rich phase is surrounded by a gradually vanishing gradient of concentration reaching deep into the surrounding matrix. Similar effect is visible in Fig. 6 in a 75 percent Mg, 25 percent Pb alloy, which is in "as cast" condition.*

Perhaps the most striking comparison of a microradiograph and a photomicrograph is in Fig. 7 which shows a 70 percent Al 30 percent Ag and a 90 percent Al, 10 percent Zn alloy, respectively. The two Al-Ag pictures show the same area and the contrast between the chaotic photograph and the highly organized x-ray picture is striking. Many details of the photograph appear to be accidental surface irregularities having no significance as far as the underlying, highly symmetric dendritic structures, as revealed by x-rays, is concerned. The white network is naturally silver. The Al-Zn alloy pictures do not show exactly the same area but a characteristic area from the same grain. (The area of the microradiograph was accidentally badly scratched after the x-ray pictures were taken.) Here the contrast between the large etched spots on the surface and the actually quite small precipitates surrounded by a zinc gradient is very striking. It is a typical example of etching

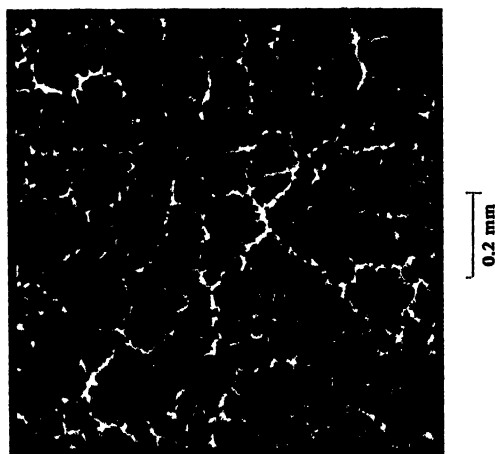


FIG. 6. Concentration gradients in a 75 percent Mg, 25 percent Pb alloy. Cu anode, filtered, 20 kv. Area: 1.1×1.1 mm.

* Solubility of Pb in Mg at 475°C is about 3.7 at. percent, at room temperature about 1.4 at. percent. The referee kindly pointed out that the blackening gradients may be caused by an oblique position of a grain boundary precipitate. Probably both factors play a role.

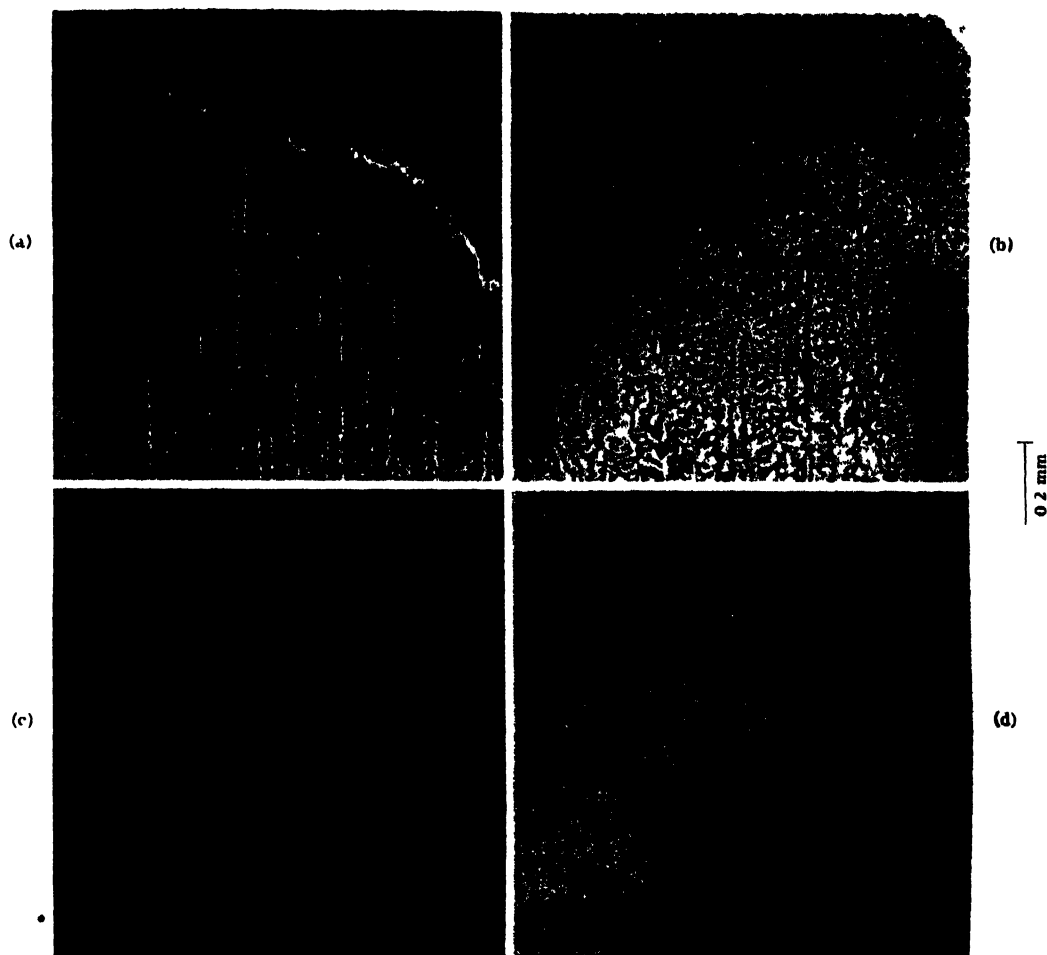


FIG. 7. Photo and x-ray microradiographs (area: 1.2×1.2 mm): (a) 70 percent Al, 30 percent Ag alloy, Mo anode, 35 kv. (b) 70 percent Al, 30 percent Ag alloy, photo. (c) 80 percent Al, 20 percent Zn alloy, Cr anode, 20 kv. (d) 80 percent Al, 20 percent Zn alloy, photo.

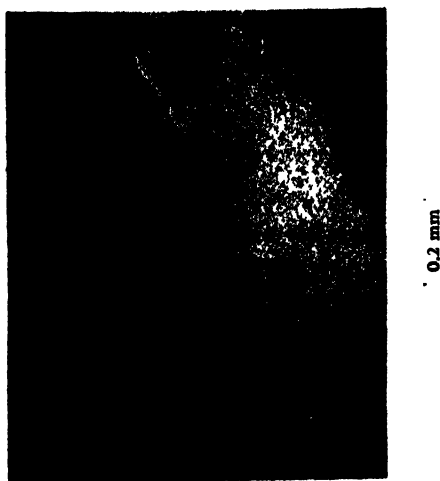


FIG. 8. Diffraction along grain boundary in an iron foil, Cr anode, 20 kv. Area: 0.8×0.9 mm.

of the "critical concentration" within the range of concentration gradient. These two examples show clearly how misleading conclusions can be drawn from conventional microscopic study of metallographic samples.

An interesting application of microradiography is the study of diffusion. One of the authors has used this method in the control of samples in a study of diffusion of copper in aluminum.⁷

B. Diffraction Effects

As discussed above, in microradiographs, there occur also diffraction effects connected with orientation of the individual grains. A few characteristic cases are illustrated in the following

⁷ F. J. Radavich and R. Smoluchowski, *Phys. Rev.* **65**, 257 (1943).



FIG. 9. Diffraction effects in a copper foil. Cu anode, filtered, 20 kv. Diameter 3.3 mm



FIG. 10 Same as Fig. 9, tilted



FIG. 11. Diffraction and concentration effects in a copper-beryllium alloy. Cu anode, filtered, 20 kv. Area: 0.9×0.7 mm

microradiographs. Figure 8 shows a typical white line parallel to the grain boundary in pure iron. Naturally, the white line gradually disappears as the boundary turns into the direction of the diffraction. The diversity and clearness of the general diffraction pattern in a polycrystalline foil is illustrated in Fig. 9 which shows a micro-radiograph of an electrolytically polished copper foil. The differences in intensity of various grains and the relative displacement of the dark and the light areas are well visible. The small spots are caused by defects in the photographic emulsion. The influence of tilting on the general diffraction pattern is illustrated in Fig. 10 which should be compared with Fig. 9. The change in

relative blackening is evident. Finally, Fig. 11 illustrates how a diffraction pattern can be differentiated from a concentration effect in the case of a copper-beryllium alloy. On tilting, the diffraction pattern due to some deformation changes while the dendritic structure remains unchanged. The diffraction from a single branch of a dendrite is particularly interesting, because it shows that even one individual member of a whole dendrite can be distorted. A more detailed microradiographic study of this alloy is under way.

These few examples indicate the correctness of the theoretical considerations and point to a new field of application of microradiography.

Transient Analysis of Linear Systems, Using Underwater Explosion Waves

M. F. M. OSBORNE* AND J. L. CARTER**
Naval Research Laboratory, Washington, D. C.

(Received June 7, 1946)

The indicial admittance of a simple mechanical system, a tourmaline crystal hydrophone, has been determined from its response to an underwater explosion wave. The indicial admittance is simply related to the observed response to the explosion wave—the observed response is just the indicial admittance “wrapped around” an exponential, $e^{-t/\tau}$. The relative frequency calibration of two hydrophones obtained by Fourier analysis of their responses to the explosion wave is in good agreement with the direct continuous wave comparison. The peak pressure of the explosion wave is also determined from this Fourier analysis. It is suggested that the explosion wave can be used as a tool in acoustical “circuit analysis” in much the same way that a square wave generator is used in electrical circuit analysis.

IN a previous paper¹ an iterative, numerical procedure was given for determining the indicial admittance, or response to the unit function, of any linear system, from its response to a known, arbitrary transient. In this paper an application of this method is made to a simple structure, a tourmaline crystal hydrophone. The transient used was the pressure wave from an underwater explosion—a No. 8 Dupont electrical blasting cap. This transient has a simple and known form,^{2,3} that of an exponential decay, $p = p_{\max} e^{-t/\tau}$. The time t is measured from the instant of arrival of the shock front, and τ is the “time constant” of decay.

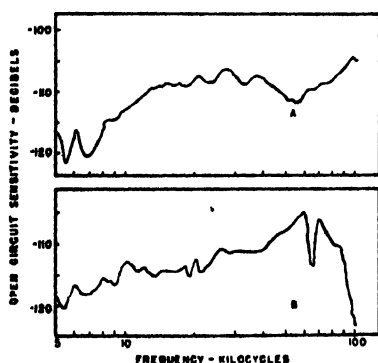


FIG. 1. Frequency calibration of hydrophones compared. 0 db = 1 volt/dyne-cm².

* Vibration Section, Sound Division.

** Mountain Lakes Laboratory, New Jersey.

¹ M. F. M. Osborne, *J. App. Phys.* **14**, 180-3 (April, 1943).

² M. F. M. Osborne and S. D. Hart, “Transmission, reflection and guiding of an exponential pulse by a steel plate in water. II. Experiment,” *J. Acous. Soc. Am.* **18**, 170 (1946).

³ M. F. M. Osborne and A. H. Taylor, “Non-linear propagation of underwater shock waves,” *Phys. Rev.* **70**, 322 (1946).

The explosion wave was also used to obtain the relative frequency calibration of two hydrophones, using methods quite similar to those used by Bedford and Fredenhall⁴ in obtaining the frequency calibration of a television amplifier with a square wave generator. Since the absolute calibrations of the hydrophones were known independently, this analysis also gave information about the explosion wave.

The two hydrophones analyzed were similar and consisted of a stack of tourmaline disks, set in a metal sphere filled with castor oil, with a rubber acoustic “window” over the face of the crystals. The frequency response curves of the two hydrophones *A* and *B* are shown in Fig. 1. Oscillograms of the responses of these hydro-

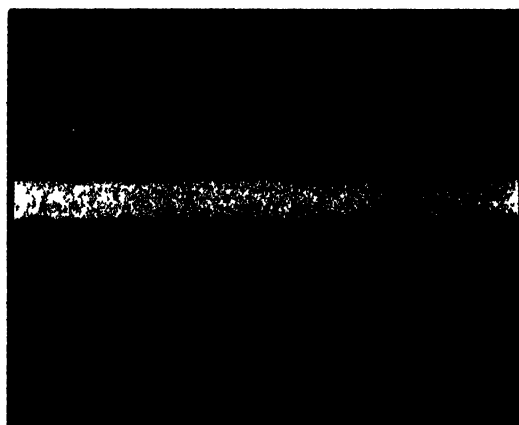


FIG. 2. Typical oscillograms of hydrophone *A*. The spot modulation frequency is 200 kilocycles/sec.

⁴ A. V. Bedford and G. L. Fredenhall, *Proc. I.R.E.* **30**, 440-447 (October, 1942).

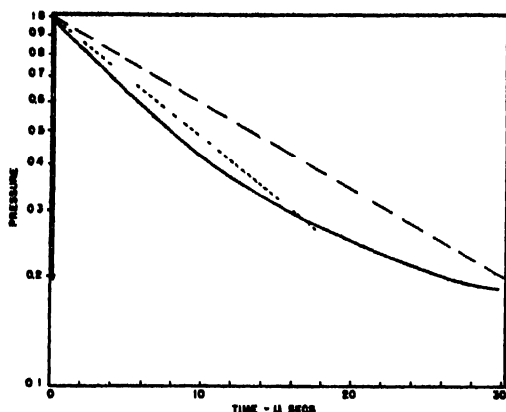


FIG. 3. The incident transient, or pressure in the explosion wave as a function of time. The ordinate units are arbitrary, being given in terms of the peak pressure as unit.

phones to the explosion wave were obtained with a television amplifier and cathode-ray oscillograph.² Typical oscillograms are shown in Fig. 2. The time scale is given by modulation of the spot intensity.

In order to obtain the indicial admittance of the hydrophone from the observed oscillograms, one has to know the true shape of the explosion wave. This in turn has to be obtained from oscillograms from other pressure gauges known to be flat over a frequency range much greater than the range of interest of the hydrophone under investigation. The adopted shape of the explosion wave, taken as a mean from a number of gauges,^{2,3} is given in Fig. 3, full curve. The dotted and dashed straight lines are exponential approximations, $p = p_{\max}e^{-t/\tau}$.

Using the method previously described,¹ with the full curve of Fig. 3 as the incident transient, the indicial admittance (full curve) of Fig. 5 was obtained. If the straight line approximations of Fig. 3 are used as the incident transient, the analysis is considerably simplified, as follows.

If $1(t)$ is the unit function, and $H(p)$ the impedance operator of the hydrophone (here $p = d/dt$, not pressure), $A(t)$ the indicial admittance, then

$$H(p) \cdot 1(t) = A(t). \quad (1)$$

Now the impedance operator or operational equivalent for a simple exponential function is $p/(ap+b)$, so that

$$p/(ap+b) \cdot 1(t) = p_{\max}e^{-t/\tau}, \quad (2)$$

where $a = 1/p_{\max}$, $b = 1/p_{\max}\tau$. (This is easily

shown from the differential equation of a circuit consisting of a battery, resistance, and capacitance in series.)

Hence the inverse operator, $a+b/p$, operating on the exponential explosion wave $p_{\max}e^{-t/\tau}$ converts it into the unit function

$$(a+b/p) \cdot p_{\max}e^{-t/\tau} = 1(t). \quad (3)$$

Now the impedance operator $H(p)$ operating on the explosion wave gives just the observed oscillogram $O(t)$

$$H(p) \cdot p_{\max}e^{-t/\tau} = O(t).$$

Hence, using (3)

$$(a+b/p)H(p)p_{\max}e^{-t/\tau} = (a+b/p) \cdot O(t),$$

$$H(p) \cdot 1(t) = A(t) = (a+b/p) \cdot O(t).$$

One has therefore the very simple rule; to convert the response to the exponential (observed oscillogram) to the indicial admittance, one simply takes a linear combination of the observed oscillogram and its indefinite integral, i.e., the ordinates of the oscillogram and its "integral curve" (Fig. 4) are combined in constant proportion. The result of this process is shown by the dotted and dashed curves of Fig. 5, corresponding to the curves of Fig. 4. They fall below the full curve, corresponding to the overestimate of the time constant (b too small).

It will be observed, by comparing Figs. 4 and 5, that the contribution of the integrated term becomes more and more nearly a constant, so that the observed oscillogram is, after a short time, just the indicial admittance lowered by a constant amount. Essentially, the observed oscillogram is just the true indicial admittance

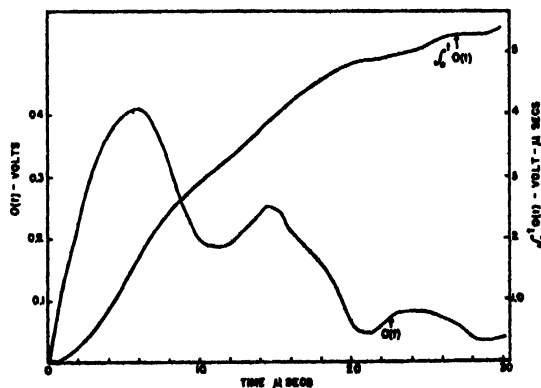


FIG. 4. The observed oscillogram for hydrophone A and its integral curve.

"wrapped around" an exponential type function. However, if the time constant of the explosion wave is *very* short compared to the characteristic periods of the structure under investigation, the observed response will be approximately the derivative of the indicial admittance.

The simple relation which exists between the indicial admittance and the response to the explosion wave suggests that the explosion wave can be used as a tool in acoustical "circuit analysis" in much the same way that a square wave generator is used in electrical circuit analysis.²

As a check on the consistency of the data (but not on the assumption that the explosion wave is an exponential) one can perform Fourier analyses of the oscillograms from two different hydrophones and compare the ratio of these with the ratio of the calibrations determined directly in continuous waves. This was done by graphical methods⁴ and also by using a mechanical harmonic analyzer whose distinguishing characteristic was its ability to analyze a curve of arbitrary base length.⁵ The result of this analysis is shown in Fig. 6 where it will be seen that the explosion wave comparison is in good agreement with the continuous wave comparison.

Since the absolute calibration of the hydrophones is known, this analysis can also give the spectrum of the explosion wave. Actually the Fourier transform of the derivative of the explosion wave is given, which is equal to the fre-

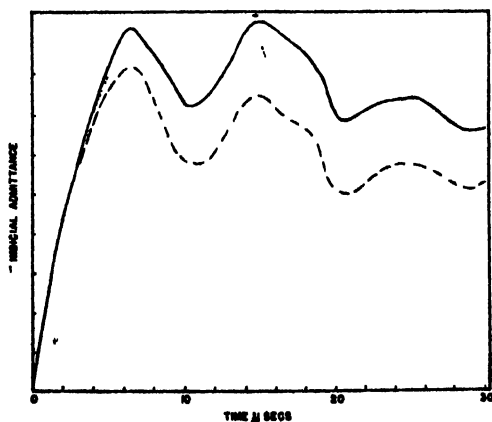


FIG. 5. The indicial admittance of hydrophone A, using Figs. 3 and 4. The ordinate scale is arbitrary, the units can be either open circuit volts/dyne-cm² or coulombs/dyne-cm².

⁵ O. Mader, *Electrotech. Zeits.* 30, 847 (September 9, 1909).

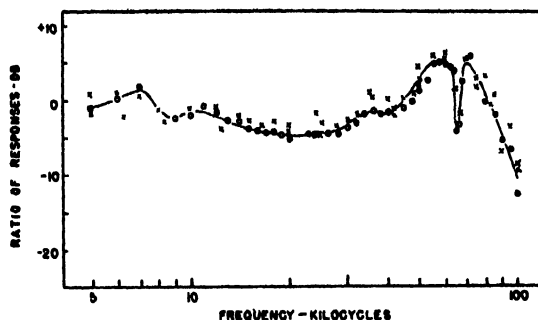


FIG. 6. Ratio of frequency calibrations of hydrophones A and B. Crosses from explosion wave data, circles from continuous wave comparison.

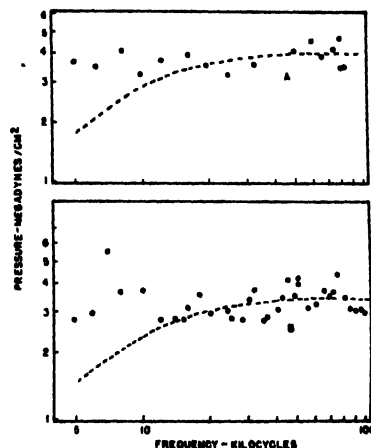


FIG. 7. "Spectrum" of explosion wave. The dotted curve is the theoretical curve for a simple exponential.

quency time the transform of the explosion wave itself. This is shown in Fig. 6. The limiting ordinate for high frequencies is the peak pressure $p_{\max}$. This is $\approx 3.5 \times 10^6$ dynes/cm² at a distance of 18 feet (5.5 meters) from the explosion, a value in good agreement with the expected value from other more direct methods.^{2,3} The dotted curve in Fig. 7 is the theoretical curve for an exponential explosion wave corresponding to the dotted curve of Fig. 5. Systematic deviations appear at the low frequency end. This end of the curve is poorly determined experimentally, since the oscillograms from which the data were obtained only lasted one or two periods of the corresponding frequency (~ 5 kc). The deviations of points from the simple theory are in the right direction (above the curve) since there are some theoretical reasons for thinking that the explosion wave dies off less rapidly than an exponential for large values of the time.

An X-Ray Method for Measuring the Thickness of Thin Crystalline Films

A. EISENSTEIN*

Radiation Laboratory,** Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 31, 1946)

When x-rays are scattered from a thin crystalline surface film overlying a crystalline base material, diffraction lines are observed from both materials. Equations are developed for the general case of (1) a flat sample and (2) a cylindrical sample expressing the ratio of line intensities from the two materials as a function of film thickness. A usable range of 10^{-8} cm to 5×10^{-2} cm is indicated. Experimental confirmation is found in the range 10^{-4} cm to 6×10^{-3} cm.

INTRODUCTION

ALTHOUGH electron diffraction methods are generally used to detect and analyze thin surface films, x-ray diffraction patterns have been obtained¹ from electrodeposited films of copper as thin as 5×10^{-6} cm. In these experiments determinations of thickness were made from the time and current flow conditions of the electrodeposition process rather than from the x-ray patterns which served only to indicate the crystal structure present. Since diffraction patterns of thin surface films are composed of diffraction lines from both the underlying base material and surface layer, it seems reasonable to expect that a comparison of the relative, integrated intensities of the two sets of diffraction lines can be used as a measure of the surface film thickness. Such a comparison must take into account differences in absorption, crystal structure, and structure factor of the two materials as well as geometrical considerations. Equations are developed for use in obtaining the thickness of thin surface films on either flat or cylindrical samples. An experimental verification is found in the case of the cylindrical sample.

Since the completion of this work two abstracts have appeared describing thickness measurements of thin coatings using x-ray techniques. One of these methods,² based essentially on absorption, compares the intensities of certain diffraction lines of the base material

observed before and after coating with the surface layer. From this measured intensity ratio, the absorption coefficient of the surface material and the geometry of the arrangement, it was possible to determine metal plating thicknesses in the range 10^{-5} cm to 10^{-2} cm. The second method,³ essentially that which is described in this paper, compares the intensities of diffraction lines from the surface and base materials. An equation is shown for the special geometrical condition of back reflection and is verified experimentally in the thickness range of 10^{-4} cm.

FLAT SAMPLE

Figure 1 shows a collimated, homogeneous, primary beam of wave-length λ and intensity I_0 falling upon a mass of crystalline powder having a total volume V . A total power P is diffracted into the Debye-Scherrer halo forming a cone of semi-apex angle 2θ . The power which is diffracted into a circular segment of length s and is recorded by a cylindrical strip of film is given by⁴

$$P' = \frac{I_0 s e^4 \lambda^3}{16 \pi r m^2 c^4} N^2 V p F^2 \frac{1 + \cos^2 2\theta}{\sin 2\theta \sin \theta} A(\theta), \quad (1)$$

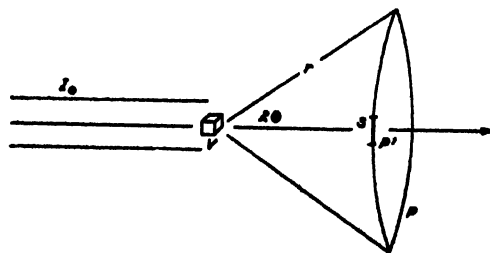


FIG. 1. Debye-Scherrer scattering from volume V .

* R. B. Gray, *Bull. Am. Phys. Soc.* 20, No. 5 (1945).

⁴ *International Tables for the Determination of Crystal Structures* (Gebrüder Borntraeger, Berlin, 1935), p. 561.

* Now at the Physics Department, Massachusetts Institute of Technology.

** This paper is based on work done for the Office of Scientific Research and Development under contract OEMar-262 with Massachusetts Institute of Technology.

¹ G. L. Clark, G. Fish, and L. E. Weeg, *J. App. Phys.* 15, 193 (1944).

² L. S. Birks and H. Friedman, *Bull. Am. Phys. Soc.* 20, No. 4 (1945).

where

N = number of unit cells per unit volume or $1/V_c$ where V_c is the unit cell volume,

p = multiplicity,

F = crystal structure factor,

r = camera radius, and

$A(\theta)$ = correction for absorption in the sample.
 e , m , and c have their usual significance.

The power dP_1' diffracted from the volume element dV_1 is,

$$dP_1' = KN_1^2 p_1 F_1^2 (\text{L.P.})_1 A(\theta) dV_1, \quad (2)$$

where

$$K = (I_0 e^4 \lambda^3) / (16 \pi r m^2 c^4), \quad (3)$$

and includes those terms which are independent of the particular volume element chosen. (L.P.) designates the usual Lorentz-Polarization factor.

The geometrical arrangement for the general case of diffraction from a double layer flat sample is shown in Fig. 2. A primary beam of intensity I_0 and cross section A_0 falls on the sample at an incident angle α irradiating a projected area A . This gives rise to a diffracted power dP_1' from the volume element dV_1 of layer 1 and to dP_2' from a similar volume element of layer 2. Radiation scattered from dV_1 must traverse a total path length in layer 1 of xy_1 where

$$xy_1 = x \left(\frac{1}{\sin \alpha} + \frac{1}{\sin (2\theta_1 - \alpha)} \right). \quad (4)$$

Introducing the absorption correction and

$$dV_1 = A dx, \quad (5)$$

$$dP_1' = KAN_1^2 p_1 F_1^2 (\text{L.P.})_1 \exp \{-\mu_1 xy_1\} dx, \quad (6)$$

where μ_1 is the linear absorption coefficient and F_1 the crystal structure factor for the material of this layer. By integration between the limits 0 and l , the power scattered from layer 1 of thickness l ,

$$P_1' = KAN_1^2 p_1 F_1^2 (\text{L.P.})_1 \frac{(1 - \exp \{-\mu_1 ly_1\})}{\mu_1 y_1}. \quad (7)$$

Consider the scattering from the volume element dV_2 of layer 2 which is located at a depth greater than l below the surface. The total path traversed by the diffracted rays is

$$ly_2 + xy_2 \quad (8)$$

in layers 1 and 2, respectively. By substitution

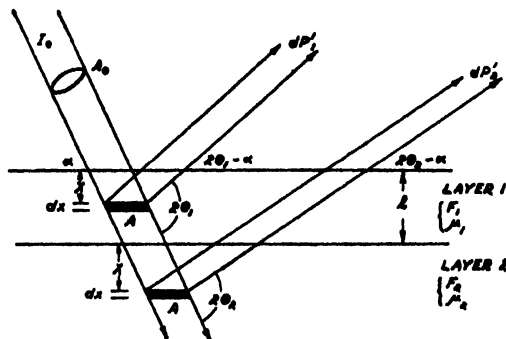


FIG. 2. Diffraction from double layer flat sample.

in (2) the diffracted power is,

$$dP_2' = KAN_2^2 p_2 F_2^2 (\text{L.P.})_2 \exp \{-\mu_1 ly_2\} \exp \{-\mu_2 xy_2\} dx. \quad (9)$$

If the absorption coefficient μ_2 is made large, either by a suitable choice of backing material or primary beam wave-length, layer 2 may be considered as having an infinite thickness for purposes of integration. Hence, integration between the limits 0 and ∞ yields,

$$P_2' = KAN_2^2 p_2 F_2^2 (\text{L.P.})_2 \frac{\exp \{-\mu_1 ly_2\}}{\mu_2 y_2}. \quad (10)$$

The ratio of integrated intensities from the materials of layers 1 and 2 is (7)/(10) or

$$\frac{P_1'}{P_2'} = \frac{N_1^2 p_1 F_1^2 \mu_2 y_2 (\text{L.P.})_1}{N_2^2 p_2 F_2^2 \mu_1 y_1 (\text{L.P.})_2} \frac{1 - \exp \{-\mu_1 ly_1\}}{\exp \{-\mu_1 ly_2\}}. \quad (11)$$

When it is convenient to select for comparison, diffraction lines from the two materials falling at approximately the same angular position, the ratio (11) may be simplified. If $\theta_1 = \theta_2$, then $y_1 = y_2$ and $(\text{L.P.})_1 = (\text{L.P.})_2$ and

$$l = \frac{1}{\mu_1 y} \ln \left\{ \frac{P_1' N_2^2 p_2 F_2^2 \mu_1}{P_2' N_1^2 p_1 F_1^2 \mu_2} + 1 \right\}. \quad (12)$$

The thickness of the surface material may thus be determined from measurements of the integrated intensities of selected diffraction lines from the two layers. A knowledge is also required of the unit cell sizes and the linear absorption coefficients of the two materials as well as the crystal structure factors and multiplicities for the particular hkl reflections.

If the materials of layers 1 and 2 have approximately the same N , p , F , and μ values, the

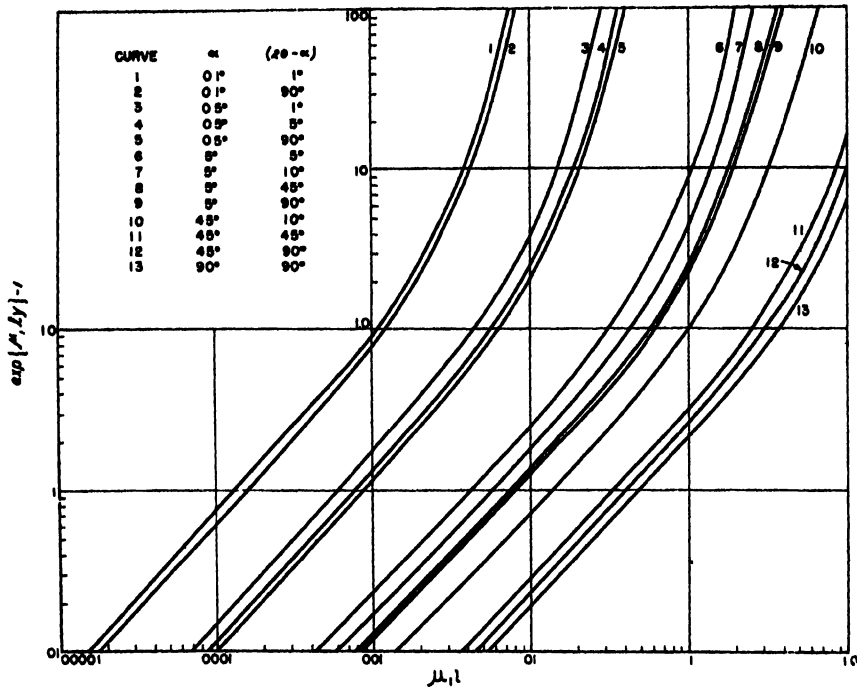


FIG. 3. $(\exp \mu_1 l y - 1)$ vs $\mu_1 l$, see text.

ratio of diffracted powers P_1'/P_2' is simply $(\exp \{\mu_1 l y\} - 1)$. A plot of this exponential function vs. $\mu_1 l$ is shown in Fig. 3 for various values of the incidence angle α and the reflection angle $(2\theta - \alpha)$. Because of the interchangeability of α and $(2\theta - \alpha)$ in this function, these curves cover the experimental range from glancing incident angle, 0.1° , to normal incidence angle, 90° , as used in the back reflection x-ray method. In general, the variation of the exponential function with the reflection angle $(2\theta - \alpha)$ is small for a given incidence angle α , hence the diffraction lines selected for comparison need only fall in the same general angular range. Assuming the ratio P_1'/P_2' can be measured between 0.01 and 100, the experimental limits of thickness measurements by this method are seen to depend critically upon the α and $(2\theta - \alpha)$ values selected.

As an example of the application of this method, consider the simple case of a surface layer which has the same crystal structure as that of the underlying material but different lattice constant, atomic structure factor f , and absorption coefficient, e.g., silver over copper. Selecting the 400 line of Ag and the 222 line of Cu which fall at about the same angular position,

$2\theta = 95^\circ$, Table I shows values to be substituted in (12). Atomic structure values as well as linear absorption coefficients are given for the radiation wave-length of $\text{CuK}\alpha$, 1.54Å. Assuming the ratio P_1'/P_2' measurable between 0.01 and 100, a substitution of these values into (12) leads to a measurable thickness range for glancing incidence, $\alpha = 0.5^\circ$, of 10^{-7} cm to 2×10^{-5} cm and for back reflection, $\alpha = 90^\circ$, of 10^{-6} cm to 2×10^{-4} cm. The lower limit of thickness measurement could be extended, in the case of glancing incidence, by the selection of lines falling at a smaller value of 2θ , e.g., 111 of Cu and 200 of Ag. Irregularities in the surface will tend, however, to increase the practical lower limits of measurement.

CYLINDRICAL SAMPLE

Figure 4 shows a cylindrical sample of radius R and length L intercepting a primary beam of collimated radiation and scattering a diffracted

TABLE I.

	N	ρ	f	$F = \rho f$	μ
Ag (400)	$1/(4.01)^3$	6	24.8	99.2	2280
Cu (222)	$1/(3.61)^3$	8	13.7	54.8	455

beam at an angle 2θ with the incident direction. When the absorption coefficient of either layer is large, only the surface areas AB and $A'B'$ contribute to the diffracted beam. The total diffracted power P'_{T_k} , falling in the segment s of Fig. 1, can be represented as the summation of scattering dP'_k from flat surface elements dA integrated over the effective scattering surface AB . Then,

$$P'_{T_k} = \int_A^B dP'_k, \quad (13)$$

where dP'_k is obtained by substituting dA for A in (7) or (10). Since the surface element

$$dA = LRd\alpha, \quad (14)$$

substitution into (13) yields,

$$P'_{T_1} = \frac{KLRN_1^2 p_1 F_1^2 (L.P.)_1}{\mu_1} \times \int_0^{2\theta_1} \frac{1 - \exp \{-\mu_1 l y_1\}}{y_1} d\alpha, \quad (15)$$

and

$$P'_{T_2} = \frac{KLRN_2^2 p_2 F_2^2 (L.P.)_2}{\mu_2} \times \int_0^{2\theta_2} \frac{\exp \{-\mu_1 l y_2\}}{y_2} d\alpha; \quad (16)$$

and the ratio (15)/(16),

$$\frac{P'_{T_1}}{P'_{T_2}} = \frac{N_1^2 p_1 F_1^2 \mu_2 (L.P.)_1}{N_2^2 p_2 F_2^2 \mu_1 (L.P.)_2} \times \frac{\int_0^{2\theta_1} \frac{1 - \exp \{-\mu_1 l y_1\}}{y_1} d\alpha}{\int_0^{2\theta_2} \frac{1 - \exp \{-\mu_1 l y_2\}}{y_2} d\alpha}. \quad (17)$$

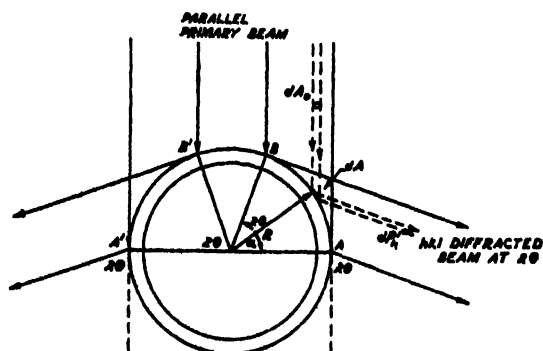


FIG. 4. Diffraction from double layer cylindrical sample.

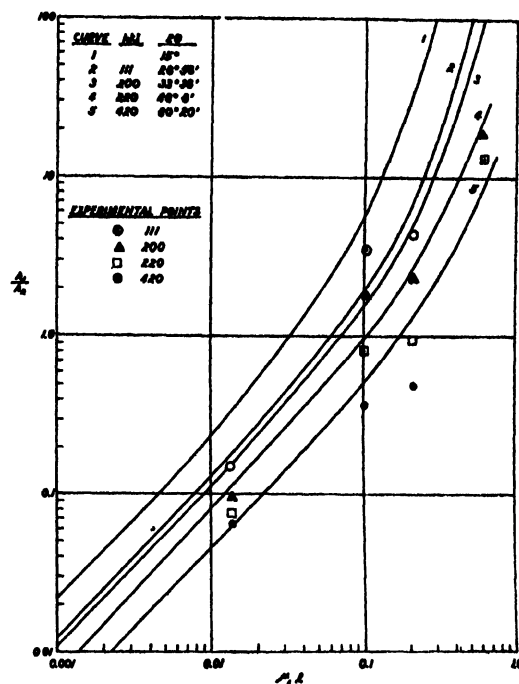


FIG. 5. Ratio of integrals of (15) and (16), A_1/A_2 , respectively, vs. $\mu_1 l$.

TABLE II.

hkl	111	200	220	420
2θ	28°51'	33°36'	48°8'	80°20'

The integrals of (15) and (16) have been solved only by graphical integration and these values are designated as A_1 and A_2 , respectively.

Selecting again for comparison, lines in the same angular range $\theta_1 = \theta_2$, then $y_1 = y_2$ and $(L.P.)_1 = (L.P.)_2$, hence

$$\frac{A_1}{A_2} = \frac{P'_{T_1} N_2^2 p_2 F_2^2 \mu_1}{P'_{T_2} N_1^2 p_1 F_1^2 \mu_2}. \quad (18)$$

The ratio A_1/A_2 was evaluated as a function of $\mu_1 l$ for specific values of 2θ . The exact values selected were 15° and those which correspond to the angular positions of diffraction lines from a (BaSr)O equal molar solid solution for Cu $K\alpha$ radiation, see Table II. The ratio A_1/A_2 is shown as a function of $\mu_1 l$ in Fig. 5. This set of curves for the case of a cylindrical sample, are seen to correspond roughly to the set in Fig. 2 for a flat sample set at an incident, glancing angle of 5°.

In practice, the ratio A_1/A_2 is obtained by means of (18) from the experimentally measured, integrated intensity ratio P'_{T1}/P'_{T2} and the thickness l determined from Fig. 5. As an example, consider again the case of an Ag layer covering a Cu base structure and compare the diffracted 400 and 222 lines, respectively. Assuming the ratio P'_{T1}/P'_{T2} can be measured between 0.01 and 100, substitution into (18) indicates a measurable thickness range of 3×10^{-8} cm to 10^{-3} cm. The selection of diffraction lines occurring at smaller 2θ positions would decrease the limits of this range, as indicated in Fig. 5.

An inspection of (7) and (10) shows that the diffracted power dP_λ per unit area dA of periphery is symmetrical with respect to variations in α , about $\alpha = \theta$. Hence, there is an asymmetrical distribution of intensity, power per unit area of diffracted beam, across the beam as it is intercepted by the film. This effect is of interest in assigning an appropriate correction for the angular displacement of diffraction lines and has been discussed⁸ in this connection.

EXPERIMENTAL

A series of cylindrical thickness standards were prepared to establish experimentally the validity of Eq. (18). The samples used were in the form of indirectly heated oxide cathodes. Nickel cylinders 3 mm in diameter were sprayed with a coating of BaCO_3 , sufficiently thick to prevent scattering from the underlying base metal. Over these were sprayed thin, uniform coatings of SrCO_3 of differing weights. By heating in vacuum at an uncorrected pyrometer temperature of 800°C , the carbonates were converted to the oxides BaO and SrO. At this temperature no interdiffusion of the two layers to form a solid solution could be detected in the

x-ray patterns of the wax protected cathodes.⁷ These patterns were obtained using filtered Cu $K\alpha$ radiation, a 3×0.5 -mm slit collimator and a 4.70-cm radius camera. Microphotometer tracings were made of the lines diffracted from the 111, 200, 220, and 420 planes of SrO and BaO. As both materials have the same crystal structure and differed by only 15 percent in lattice constant, the hkl lines of SrO were compared directly with the corresponding lines of BaO which fell in approximately the same angular range. The ratio P'_{T1}/P'_{T2} was simply the ratio of maximum line intensities I_1/I_2 since the shape of all diffraction lines appeared to be the same. From the spray conditions, the apparent density of the BaCO_3 and SrCO_3 layers was known to be 1.2 ± 0.2 g/cm³ and 0.8 ± 0.2 g/cm³, respectively. SrO coating thicknesses were computed from the apparent SrCO_3 density, the weight of each coating, and the coated area. These were found to be 1.3×10^{-4} cm, 1.0×10^{-3} cm, 2.0×10^{-3} cm, and 6.3×10^{-3} cm. No correction was made for the small reduction in thickness and increase in density produced in both layers during the conversion process. The experimental points, obtained from (18) with $p_1 = p_2$ and appropriate N , F , and μ values, for the oxides using the apparent density values, fall on the $A1/A2$ curves as indicated in Fig. 5.

This method of measuring film thickness has been used to determine, (1) the thickness of electrodeposited layers, (2) the extent of surface oxidation after the exposure of the sample to an oxidizing atmosphere, (3) the thickness of sprayed oxide coatings, and (4) the thickness of surface layers produced by a chemical reaction.

The author is indebted to Professor B. E. Warren for his very helpful criticism of this manuscript as well as to Miss Eleanor Uhl for the evaluation of the function presented in Fig. 5.

⁷ A. Taylor and H. Sinclair, Proc. Phys. Soc. 57, 108 (1945).

⁸ B. E. Warren, J. App. Phys. 16, 614 (1945).

⁷ A. Eisenstein, J. App. Phys. 17, 434 (1946).

A New Bridge Photo-Cell Employing a Photo-Conductive Effect in Silicon. Some Properties of High Purity Silicon*¹

G. K. TEAL, J. R. FISHER, AND A. W. TREPTOW
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received July 2, 1946)

A pure photo-conductive effect was found in pyrolytically deposited and vaporized silicon films. An apparatus is described for making bridge type photo-cells by reaction of silicon tetrachloride and hydrogen gases at ceramic or quartz surfaces at high temperatures. The maximum photo-sensitivity occurs at 8400–8600Å with considerable response in the visible region of the spectrum. The sensitivity of the cell appears about equivalent to that of the selenium bridge and its stability and speed of response are far better. For pyrolytic films on porcelain there are three distinct regions in the conductivity as a function of temperature. At low temperatures the electronic conductivity is given by the expression $\sigma = Af(T) \exp-(E/2kT)$. At tempera-

tures between 227°C and a higher temperature of 400–500°C $\sigma = A \exp-(E/2kT)$, where E lies between 0.3 and 0.8 ev; and at high temperatures $\sigma = A \exp-(E/2kT)$, where $E = 1.12$ ev. The value 1.12 ev represents the separation of the conducting and non-conducting bands in silicon. The long wave limit of the optical absorption of silicon was found to lie at approximately 10,500Å (1.18 ev). The data lead to the conclusion that the same electron bands are concerned in the photoelectric, optical, and thermal processes and that the low values of specific conductances found (1.8×10^{-8} ohm⁻¹ cm⁻¹) are caused by the high purity of the silicon rather than by its polycrystalline structure.

I. INTRODUCTION

IT has long been known that the electrical conductivity of certain solid insulating or semi-conducting materials can be temporarily increased by illumination with light of suitable wave-length.² Several such materials have been employed in resistance bridge type of photo-cells and have been commercially available. The conductivity imparted by light is electronic in character and results from a photoelectric effect occurring within the individual crystals. Two types of crystals have been recognized in studies of the photo-conductivity of various solids. Some materials exhibit photo-conductivity when chemically pure and in the form of perfect or nearly perfect single crystals. These are known as idiochromatic crystals; typical examples are diamond, zincblende, sulphur, the red variety of selenium, stibnite, and cinnabar. Other materials show no photo-conductivity in the pure state and acquire it only through incorporation

of impurities into the regular crystal lattice, by irradiation by x-rays or ultraviolet light, by application of high electric fields, or by thermal treatment to displace constituents from their normal lattice positions or partially to decompose the material so as to result in a stoichiometric unbalance of the normal lattice constituents. Such materials are known as allochromatic crystals and include the alkali, silver, and thallium halides. There are, in addition to the particular materials mentioned, a great number of other materials that have been investigated in polycrystalline form and found to be photo-conducting but which have not been studied sufficiently to permit their classification.

The process of conduction has been shown to be the sum of a primary effect and secondary effects. The primary process is instantaneous whereas the secondary process is the result of the passage of current and may take many minutes before reaching an equilibrium state. It is of considerable practical import for some purposes, therefore, that the material employed in a resistance bridge photo-cell shall have a highly sensitive primary current accompanied by little or no secondary current.

The best known commercial bridge photo-cells are perhaps the selenium and thalofide cells. The photo-conductive effect in the selenium cell is accompanied by large secondary effects. The

* This paper is based in whole or in part on work done for the Office of Scientific Research and Development under Contract OEMar-1231 with the Western Electric Company.

¹ Presented at meeting of the Am. Phys. Soc. April 26, 1946. Phys. Rev. 69, 686A (1946). See also NDRC Sec. 16.4-37 Report, Jan. 4, 1945, G. K. Teal.

² For summary see A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena* (McGraw-Hill Book Company, Inc., New York, 1932), Chap. VIII; F. C. Nix, Rev. Mod. Phys. 4, 723 (1932); A. L. Hughes, Rev. Mod. Phys. 8, 294 (1936).

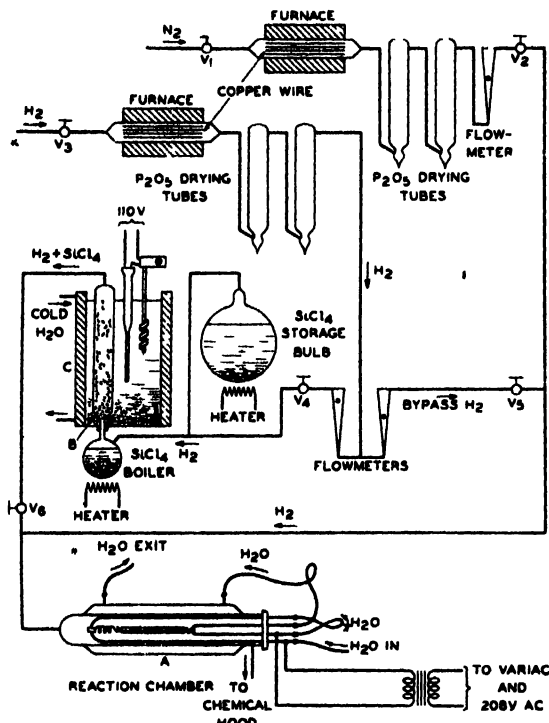


FIG. 1.

secondary effects in thalofide, however, do not appear to be so serious.

During the war several investigators observed photo-effects in silicon. In the present paper we shall describe a pure photo-conductivity found in silicon³ and a method of preparing bridge photo-cells employing the effect. Experimental evidence will be presented tending to prove that silicon is an idiochromatic crystal with minor secondary effects. Several attractive features of the cells will be described.

II. METHODS OF PREPARING CELLS

Silicon cells may be prepared by use of pyrolytically deposited or vaporized silicon. The cells described below were made by pyrolytic deposition of silicon on ceramic and quartz bases. For this purpose the reduction of silicon tetrachloride vapors by hydrogen gas at a hot surface⁴ is satisfactory. It seems likely that this

³ Since preparing the present paper two reports on other photoeffects in silicon have come to our attention. P. H. Miller, Jr. and M. H. Greenblatt, U. of Pennsylvania OSRD Report No. 412, March 20, 1945; and F. C. Brown, *Phys. Rev.* **69**, 686A (1946).

⁴ R. Hölbling, *Zeits. f. angewandte Chemie* **40**, 655 (1927); J. N. Pring and W. Fielding, *J. Chem. Soc.* **95**, 1497 (1909).

reaction is a complicated one involving the formation of intermediate products of long chain silicon chlorides.⁵ For the purposes of simplicity, however, we shall write it as:



One method of preparing cells has been to deposit the silicon films on ceramic tubes suspended on internal tantalum wire heaters in flowing silicon tetrachloride and hydrogen gases. A schematic diagram of the experimental apparatus is shown in Fig. 1. It consists of a gas flow system for providing controlled amounts of hydrogen, nitrogen, and silicon tetrachloride to the water cooled reaction chamber A in which the ceramic tube is suspended by means of the internal wire heater, the ends of which are connected to water cooled copper electrical leads. As can be seen in the diagram the purified hydrogen may be introduced into the reaction chamber through a by-pass and also through a silicon tetrachloride supply branch. Each branch is provided with a floating ball flow-meter. The amount of silicon tetrachloride introduced into the reaction chamber is controlled by the hydrogen flow through the silicon tetrachloride boiler and also the temperature of the reflux condenser B, which is surrounded by the hollow-walled metal thermostat C. Cells have also been made by heating the ceramic or quartz base in a furnace tube connected to the gas supply system in place of the reaction chamber A. The temperature of the ceramic base is determined by means of an optical pyrometer when the internal wire heater is used and by means of a thermocouple when the furnace is employed.

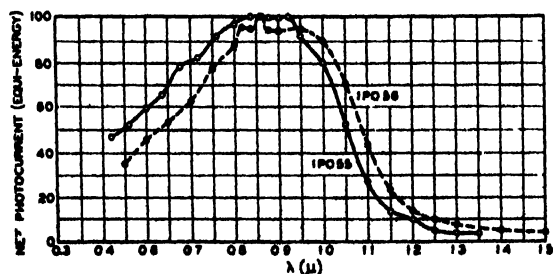


FIG. 2. Spectral sensitivity of silicon bridge photo-cells.

⁵ R. Schwartz and H. Meckback, *Zeits. f. anorg. u. allgem. Chemie* **232**, 241 (1937); R. Schwartz and R. Thiel, *ibid.* **235**, 247 (1938).

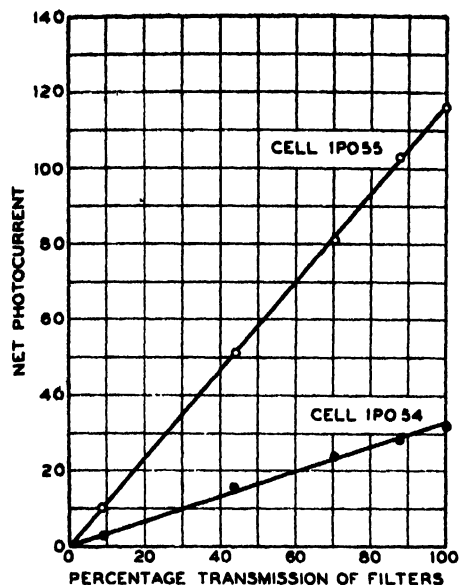


FIG. 3. Dependence of the photo-response to $\lambda 9200$ on intensity of illumination.

One type of cell consists of a thin film of silicon on the outside of a ceramic rod about 95 mils in diameter and about $\frac{3}{4}$ inch long. Another is a thin film of silicon on the flat surface of quartz or ceramic of several square inches area. The electrodes may be silver paste fired on at 600–800°C or a conducting silver paint applied and air dried at low temperatures.

Using the means just described silicon deposits readily on ceramic surfaces heated to about 1000°C or higher. The deposits made at the lower temperatures tend to be grainier than those made at the higher temperatures. In many cases the deposit made at the lower temperatures appears to the eye as a brown fuzz. Under the microscope this fuzz is seen to be long hairlike single crystals. At temperatures of about 1100°C and above the silicon deposits as a smooth, hard, gray film adhering very firmly to the ceramic base. A wide range of film thicknesses have been made. In the case of a typical film 5 microns thick a deposition time of 1.5 minutes suffices. Under a given set of gas flow conditions the rate of deposition at 1250°C has been observed to be about 5 times greater than the rate at 1000°C. The film thickness is a linear function of the deposition time. Rate of deposition is approximately proportional to the silicon tetrachloride

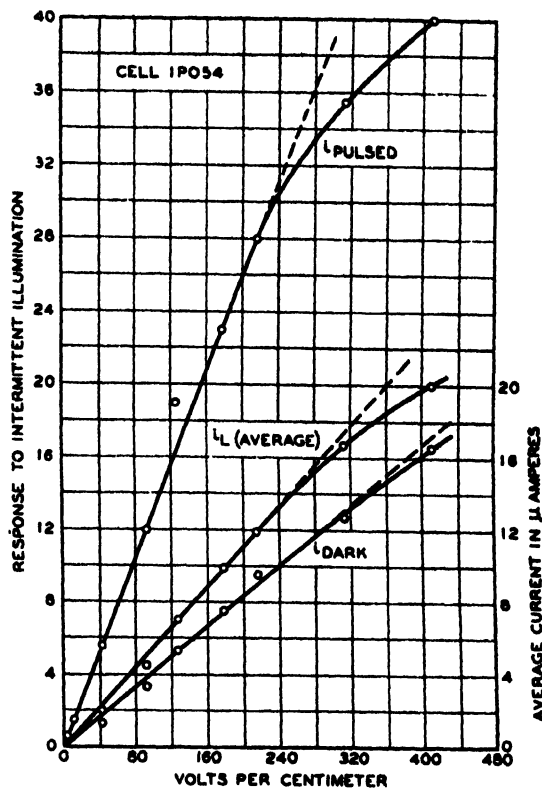


FIG. 4. Linearity of response with applied voltage gradient.

concentration up to a certain concentration above which the rate decreases.

III. CHARACTERISTICS OF THE PHOTO-CONDUCTIVE EFFECT

In Fig. 2 are shown spectral sensitivity curves of two typical cells. The maximum for each cell is set equal to 100 on an arbitrary current scale since the absolute magnitude of the photo-effect in individual cells has not been determined. However, for each separate cell the responses to light of different wave-length are on an equal energy basis. The response curves are seen to correspond closely, the maximum response in all cases lying between 8400 to 8600Å. The long wave-length limits were not determined with great accuracy but appear to lie between 12,000 to 15,000Å. The photo-sensitivity in all cases extends well into the visible region and decreases steadily as the wave-length decreases.

In order to be certain that the broadness of the spectral maxima was not due to saturation effects, the photo-conductive responses of several

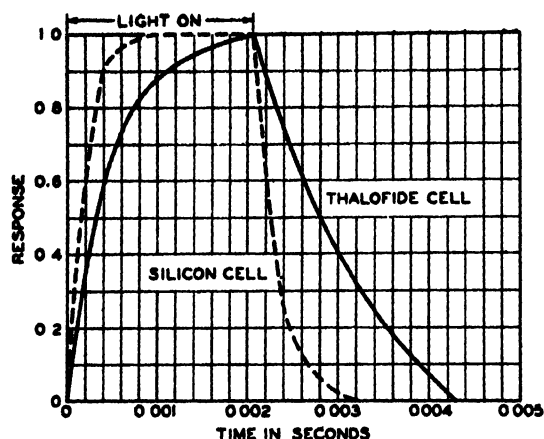


FIG. 5. Dynamic responses of silicon and thalofide cells to intermittent illumination.

cells were determined as a function of the light intensity for $\lambda = 9200\text{\AA}$ under the same conditions used in obtaining the spectral sensitivity curves. With tungsten light having a color temperature of 2848°K the maximum photo-current (uncorrected for equal energy) was obtained at about $\lambda = 9200\text{\AA}$. Levy line screens were used as filters and their transmissions were determined in the same optical setup that was used in making the measurements on the silicon cells. The results are shown, in Fig. 3, for two cells. In each case the photo-current is a linear function of the light intensity. The great breadth of the sensitivity curves near the maximum is therefore not attributable to saturation effects depressing the curve at the maximum.

In Fig. 4 there are shown plots of the response of one of the rod shaped cells to intermittent illumination as observed on a cathode-ray oscillograph. The light was interrupted by a light chopper at the rate of about 240 cycles per second. The cell was coupled to the oscillograph through a 10,000-ohm resistor. In addition the average dark and light currents were measured by means of a meter. The curves show that the a.c. response is linear with the field applied up to about 200 volts per cm. Measurements on other cells likewise indicated linearity of response with increasing field applied to the film.

The sensitivities of several rod shaped cells to uniform illumination are given in Table I. It was found that such a condition could be approximated by placing the rod cells at the center of an

equilateral triangle, at whose corners are situated three frosted bulb lamps. 40-watt bulbs were used on a regulated voltage supply. The illumination was determined by substituting a Weston photometer for the silicon cells and averaging for the three light sources. The illumination was thus found to be 127 footcandles. The potential across the cells was 48 volts in all cases. The sensitivity observed when the area illuminated is small in comparison to the diameter and length of the rod is somewhat lower than the above values and shows some variation with position of the area on the rod. The lower values, however, appear not to be associated with saturation effects due to large flux densities since observations show that the response varies linearly with the total light intensity for small spots and for the highest flux densities used. The highest sensitivities have been obtained for cells made in the glass reaction chamber of Fig. 1. The sensitivity of the silicon cells appears considerably lower than that obtainable in the thalofide cell but is considerably greater than that of a Luxtron selenium cell that we have had on hand for several years.

Some idea of the relative speeds of response of the silicon and thalofide cells can be gained from the approximate curves shown in Fig. 5. The curves were obtained on a cathode-ray oscillograph while illuminating the cells with an intermittent light signal of about 240 cycles per second. The scales are arbitrary. Though the thalofide cell shows a larger response than the silicon cell the speed of response is seen to be definitely faster in the case of silicon. The time of attainment of an equilibrium state on starting or stopping the illumination of a silicon cell is less than 0.001 second under the conditions used. In selenium the equilibrium state may not be reached for many minutes. Measurements of the response of some of the silicon cells to an

TABLE I. Sensitivity of silicon cells with uniform illumination.

Cell No.	Microamperes per lumen per 200 volts per cm	Percent change in dark current per lumen	s_i/s_d per lumen	s_i/s_d for 127 ft. candles
1P055	55.8	1720	18.2	2.007
20518F ₁	74.0	1890	19.9	2.054
20518F ₂	34.0	1040	11.4	2.193

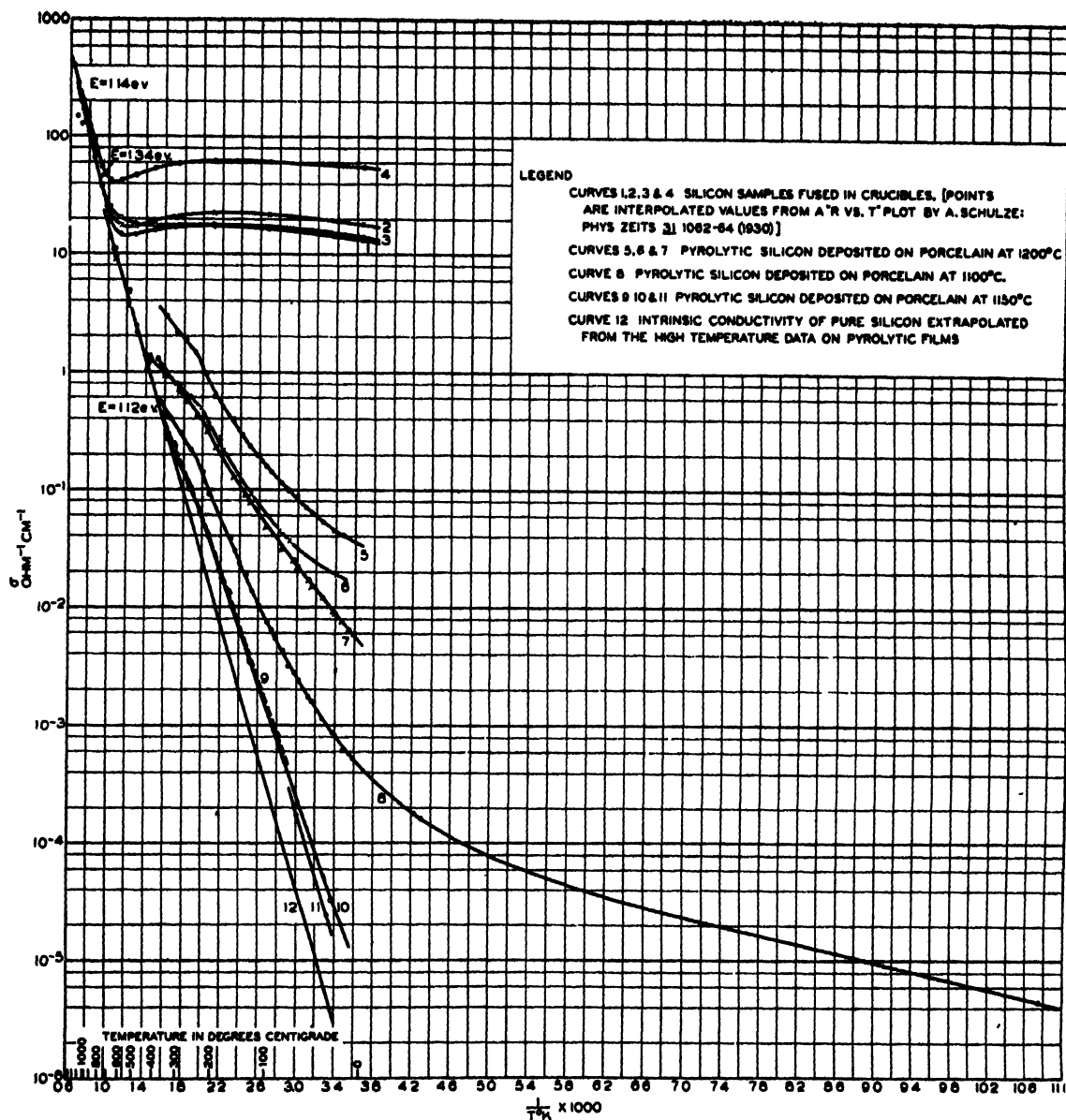


FIG. 6. Plot of the logarithm of the specific conductivity of silicon films *versus* $1/T^{\circ}\text{K}$.

intermittent light signal showed that the response at 720 cycles is 53 to 68 percent of the response at 90 cycles.⁶ For one type of thalofide cell the 720-cycle response drops to about 14 percent of the 90-cycle response.⁶

Observations⁶ of the effect of background radiation on the silicon cell indicate that it is far superior to the thalofide cell in this respect. Thus for one silicon cell studied the response to

⁶ We are indebted to Drs. W. L. Hole and L. R. Posey, Jr., of the University of Michigan for these observations.

intermittent illumination was down to 72 percent for 30 footcandles background and to 25 percent for 500 footcandles. For the thalofide cell the responsivity is down to 0.1 percent for 500-footcandles background illumination.

Silicon cells heated to a red heat in an oxygen-gas flame showed no decrease in sensitivity on cooling to room temperature. No other cell that we know of could stand such treatment. Silicon cells have been left in strong sunlight for one hour without noticeably deteriorating the response.

IV. OTHER ELECTRICAL PROPERTIES OF HIGH PURITY SILICON

Relatively few years ago the sign of the temperature coefficient of resistance of silicon was in dispute. One view was that the negative temperature coefficient often observed must be caused by thin films or other effects occurring at the intergranular boundaries of the polycrystalline material. From work done in recent years, however, it is generally accepted that silicon is a semi-conductor and that in the pure state it has a negative temperature coefficient of resistance. That a dispute arose is not surprising in view of the difficulty of preparing very pure silicon in compact form, the scarcity of experimental data on the quite enormous effects that even a few thousandths of a percent of an impurity can have on the electrical properties of a semiconductor, and also the fact that the modern theories of the solid state had not then been developed.

In Fig. 6 are shown experimental data on the effect of temperature on the electrical conductivity of a number of samples of silicon. The logarithm of the specific conductivity is plotted as a function of the reciprocal of the absolute temperature. Curves 1, 2, 3, and 4 are plotted from data on fused silicon obtained by Schulze.⁷ They are of the same general shape and give about the same magnitudes of specific conductivity obtained in more recent samples of fused silicon. The interpretation of these curves is evidently the same as that given by Seitz⁸ and Lark-Horowitz and Johnson⁹ in explaining similar but more recent data on silicon and germanium. Thus the conductivity at temperatures lower than about 500°C is primarily caused by electron transitions between impurity centers and adjacent bands while the conductivity at temperatures greater than about 600°C is primarily caused by electron transitions from the lower filled band to the higher unfilled band, and the maximum and minimum arise due to the inability of the impurities and the filled band to supply conductors fast enough to make

up for the decrease in conductivity due to lattice scattering. At high temperatures, as may be seen in the figure, the points fall on a straight line in the $\log \sigma$ versus $1/T$ plot. The slope of the line gives the energy separation E of the filled and unfilled bands in the solid according to the following equation:

$$\sigma = A \exp \left[- (E/2kT) \right],$$

where k is Boltzmann's constant, T is the absolute temperature, and A is a factor that varies rather slowly with temperature. Curves 1 and 2 coincide very well at high temperatures and lead to a value of 1.34 electron volts for the band separation. Curve 4 gives a value of 1.14 ev, which is in better agreement with Seitz's value of 1.12 electron volts. Curve 3 agrees somewhat better with the higher value from 1 and 2.

The silicon films prepared in the present work show entirely different orders of conductivity from fused silicon samples. This is shown in curves 5, 6, 7, 8, 9, 10, and 11 taken on various silicon films deposited pyrolytically on porcelain. The curves show three distinctly different regions: A straight line, high temperature region, a straight line intermediate temperature region, and a curved, low temperature region which appears to approach a constant slope at -100 to -180°C . The curvature is the reverse of that observed for fused silicon. There are thus two points of change one at about 227°C and another at 400° or higher, the exact temperature, in the latter case, depending on the magnitude of the conductivity. The points of change might be correlated with those previously indicated in the work of Hölbling¹⁰ (260° , 410° , 550°) and the discontinuous and reversible changes observed at 210° and 430°C in fused silicon by Königsberger and Schilling.¹¹ The general shape of the curves is similar to an approximate result on a sample of silicon on carbon prepared by Hölbling but the conductivities in the present work are very much lower. Also the $\log \sigma$ vs. $1/T$ plot given here makes clear that there are three regions involved and indicates why some variation may occur in the temperature at which the upper inflection point is observed. Königsberger

⁷ A. Schulze, *Physik. Zeits.* 31, 1062-4 (1930).

⁸ F. Seitz, OSRD report. PB2619, 28 pp. (1942). Office of the Publication Board of the Dept. of Commerce.

⁹ K. Lark-Horowitz and V. A. Johnston, *Phys. Rev.* 69, 158A (1946).

¹⁰ R. Hölbling, reference 4.

¹¹ J. Königsberger and K. Schilling, *Ann. d. Physik* 32, 179 (1910).

and Schilling ascribed the changes they noted to polymorphic transitions. However, such polymorphic changes have not been reported in the x-ray literature and Schulze⁷ was unable to find evidence for such changes by investigation of a number of fused silicon samples up to 1000°C with an optical dilatometer. It therefore seems probable that the changes are to be correlated with changes in the conduction mechanism.

Below 227°C the curves have a changing slope, and A in the above equation must be replaced by $Af(T)$.¹² At the lowest temperature studied the curves appear to approach a constant slope equivalent to an activation energy of about 0.08 ev. It was found also that the data of curves 5, 6, 8, 9, and 10 fit very well an empirical equation of the type $\sigma = \sigma_0 e^{\frac{T}{T_0}}$ such that a $\log \sigma$ versus T plot gives points approximating closely a straight line. In the range of very low temperatures, 100–200°K, which have received only minor study, some deviation from the straight line may occur. A somewhat more reasonable relationship might be found in the fact that $\log \sigma$ versus $\log T$ plots of the data of curves 6, 7, 8, 9, and 10 give good fits to straight lines between about 280–500°K. Some significance might attach to the fact that these lines appear to converge on a point at about 665°K. In the case of curve 5, however, this does not hold and the departure from the straight line occurs at about 380°K, such that the conductivity appears higher than would occur due to maintenance of the straight line relationship. We have no adequate interpretation of these results at the present time.

The values of activation energy for the intermediate region 227° to 500°C vary, individual values of 0.31, 0.45, 0.49, and 0.79 electron volts being obtained.

In the high temperature region individual values of E obtained are 1.13 ev, 1.106 ev, and 1.12 ev. The straight line 12 is the best fit that can be drawn for all the high temperature measurements. It represents a band separation in silicon of 1.12 electron volts, in good agreement with Seitz and with curve 4 from Schulze's work. The values of conductance are in better agreement with the data of Seitz than that of

Schulze. The high temperature data lead to a calculated value of $3.9 \times 10^{-8} \text{ ohm}^{-1} \text{ cm}^{-1}$ for the specific conductivity of pure silicon at 25°C. The lowest specific conductivity observed in the present work was $1.8 \times 10^{-8} \text{ ohm}^{-1} \text{ cm}^{-1}$. For this sample the activation energy of 0.97 electron volt was the highest value observed at room temperature.

V. SUMMARY

The best photo-conducting cells are the ones showing the lowest specific conductivity, corresponding to very pure silicon. The long wave limit of the photo-conductive effect was not measured with great accuracy but appears to lie between 12,000 to 15,000Å. The thermal activation energy of 1.12 electron volts corresponds to a $\lambda_0 = 11,000\text{Å}$. This is considered strong evidence for the same electron bands being concerned in both the photoelectric and thermal processes of increasing the conductance. The maximum photoelectric sensitivity occurs at 8400–8600Å, equivalent to 1.45 electron volts energy. It seems certain that the process is not simply a transition between an impurity level and adjacent band since this would result in a much lower value of λ_0 (in ev) than was observed.

Observations on evaporated films of silicon indicated a low value of optical absorption in the infra-red region. This is in agreement with expectations.¹³ Appreciable absorption in the shorter wave-lengths starts at about 10,500Å, increasing rapidly with decreasing wave-length until a relatively high value is attained in the neighborhood of 8750Å. The high absorption continues into the visible part of the spectrum. The approximate value 10,500Å for the long wave limit of the absorption, corresponds to a separation of the conducting and non-conducting bands of silicon equivalent to 1.18 electron volts. Such evaporated films also show photo-conductance. The maximum photoelectric sensitivity of pyrolytic films is attained at about the same wave-length at which the optical absorption becomes high.

The close agreement between the high temperature activation energies and conductance observed in this work and in the work of others

¹² F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), p. 191.

¹³ F. Moglick and R. Rompe, *Zeits. f. Physik* 119, 472 (1942). J. F. Mullaney, *Phys. Rev.* 66, 326 (1944).

on fused silicon indicates that the high resistance values and high temperature coefficients of resistance observed in the films deposited on porcelain are caused by the very high purity of the silicon and not by the crystal size nor by intergranular films. The close agreement between the activation energies obtained from the photoelectric, thermal-conductance, and optical data

suggests the phenomena have a common origin—electron transitions from the upper filled band to the conducting band.

Acknowledgment: We are very grateful to Mr. E. F. Kingsbury for assistance in determining the sensitivity of some of the cells and in making observations on the optical absorption of silicon.

The Production of High Centrifugal Fields*

J. W. BEAMS, J. L. YOUNG, III, AND J. W. MOORE

Rouss Physical Laboratory, University of Virginia, Charlottesville, Virginia

(Received July 15, 1946)

High centrifugal fields were produced by spinning small solid steel spherical rotors up to their bursting speeds. The rotors were supported magnetically in a vacuum by an improved method and spun by rotating magnetic fields. The peripheral velocities at which the rotors of various sizes, made of the same (flaw free) steel and having the same shapes, exploded, were roughly the same and of the order of 10^6 cm/sec. A centrifugal field of 2.4×10^8 times gravity was obtained with a .795-mm spherical rotor which was the smallest diameter tried. Calculations indicate that in some cases plastic flow probably occurred in small regions near the centers of the spherical rotors somewhat below their bursting speed.

WHEN substances are placed in a centrifugal field two general types of phenomena may occur. The first type depends upon $(2\pi N)^2 r^2 = v^2$ where N is the number of revolutions per sec., r is the radius of the rotor, and v is the peripheral speed or upon the energy difference between axis and periphery. The second type depends upon $(2\pi N)^2 r$ or the intensity of the centrifugal field. In this paper a method is given for producing very high centrifugal fields for use in the study of the second type of phenomena.

In a spinning homogenous elastic rotor the maximum stress produced is proportional to $(2\pi N)^2 r^2$ or the square of the peripheral speed.¹ That is, similarly shaped rotors made of the same material will explode when their peripheral velocities are approximately equal. Consequently for the production of the highest centrifugal fields which depend upon $(2\pi N)^2 r$ the radius of the rotor must be as small as possible.

For rotors made of a given elastic material the highest peripheral speed should be obtained when the shape is such that the radial and

tangential stresses are of constant value throughout the rotor.¹ If hardened high strength steel is used for the rotor, some considerable pains are

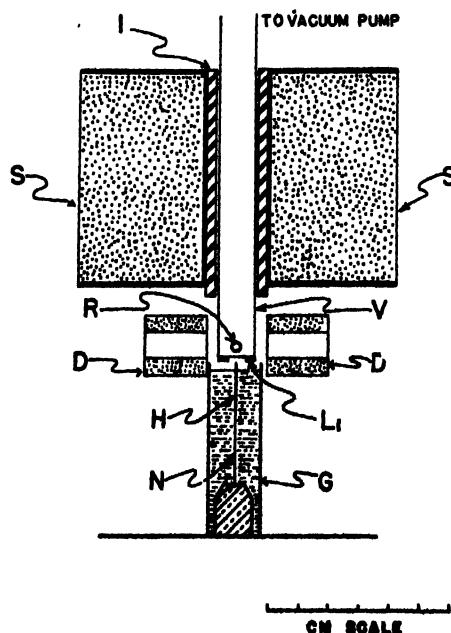


FIG. 1.

* The work described in this paper has been supported by the Bureau of Ordnance, U. S. Navy under Contract NOrd-7873.

¹ Stodola, *Steam and Gas Turbines* (McGraw-Hill Book Company, Inc., New York), Vol. I, p. 376.

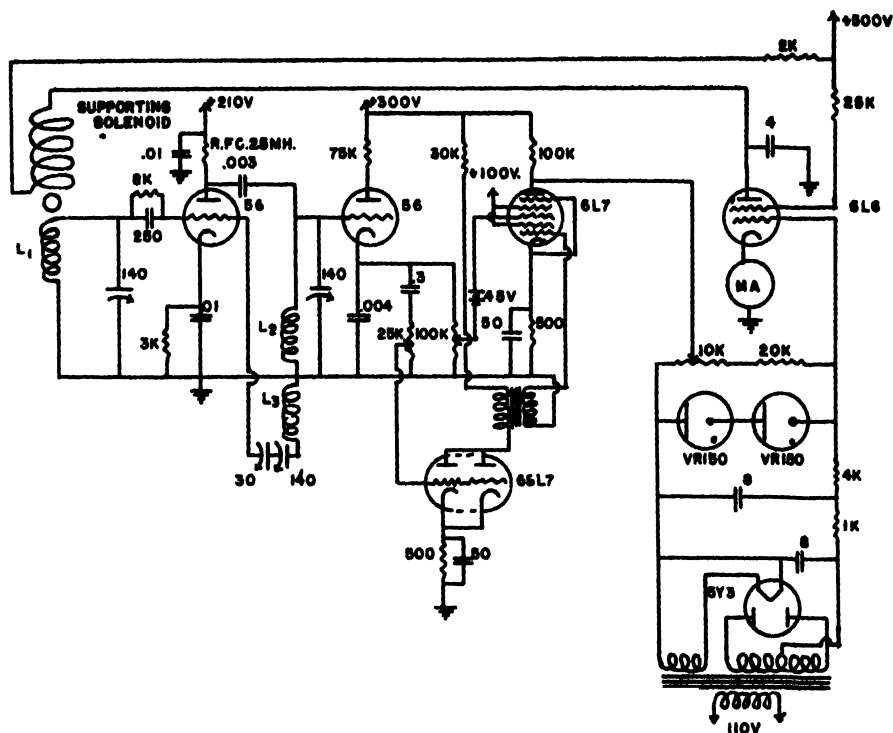


FIG. 2.

required to grind the rotor to the shape required by the above condition. However, spherical rotors are commercially available in the form of high strength solid steel balls for use in ball bearings. Consequently, these approximately spherical rotors were used in the experiments. Also in addition to their availability the calculated critical vibration frequencies of these spherical rotors are well above their bursting speeds.²

EXPERIMENTAL ARRANGEMENT

The rotors were spun by the general method developed at Virginia by a number of different workers³⁻⁶ for spinning small rotors. However, in the present experiments, considerable improvements in the apparatus have been made. Figure 1 shows schematically the arrangement used. The rotor R is suspended in the axial magnetic field of the solenoid S and is spun by a two phase rotating magnetic field in the two

pairs of coils D . The horizontal position of the steel rotor is maintained by the symmetrically diverging field of the solenoid while its vertical position is maintained by an automatic regulation of the current through the solenoid S . The small coil L_1 is part of the grid circuit of a tuned grid tuned plate radiofrequency oscillator (Fig. 2) which regulates the current through S . It is so arranged that when the rotor rises the current through S decreases. These current changes are such as to position accurately the rotor without observable hunting. An iron tube I with $\frac{3}{4}$ " O. D. $\frac{3}{32}$ " wall thickness and $3\frac{1}{2}$ " long was placed inside the solenoid to increase the field at R . A small iron wire H mounted in a glass tube G filled with a liquid assisted in damping any horizontal motion of the rotor. The glass chamber V surrounding the rotor R was evacuated by the usual fore pump, diffusion pump cold trap combination. The pressure surrounding the rotor was measured by an ionization gauge.

SUPPORTING CIRCUIT

The circuit used for supporting the rotors is shown in Fig. 2. It will be observed that the

² Love, *Mathematical Theory of Elasticity* (Dover Publications, New York), fourth edition, p. 285.

³ F. T. Holmes, *Rev. Sci. Inst.* 8, 444 (1937).

⁴ F. T. Holmes and J. W. Beams, *Nature* 140, 30 (1937).

⁵ C. S. Smith, *Rev. Sci. Inst.* 12, 15 (1941).

⁶ L. E. MacHattie, *Rev. Sci. Inst.* 12, 429 (1941).



tional electrical "damping" was introduced into the circuit. This "damping" is obtained by mixing the amplified derivative of the "error" signal with the "error" signal in the 6L7 mixer tube. The derivative of the "error" signal is obtained from the capacity-resistance circuit in parallel across the cathode follower as shown in Fig. 2. The magnitude of the capacity and of the resistance in this circuit was obtained by observing the hunting frequency of the rotor and then making the capacitative reactance and resistance such that the derivative signal would not be too small and at the same time more than 80° out of phase with the variation of the "error" voltage. The values are given in Fig. 2. This derivative signal is amplified by the two triodes of a 6SL7 in parallel and applied through transformer coupling to the second control grid of the 6L7 mixer tube. This arrangement greatly stabilized the rotor and reduced all vertical hunting and oscillations to such an extent that no vertical motion of the rotor could be observed with a 20-power microscope focused on markings of the rotor. The values of the resistances, capacities and inductances are given in Fig. 2. The supporting solenoid *S* was wound with 20,500 turns of No. 30 wire. Its inside diameter

was $\frac{1}{8}$ " and its length $2\frac{1}{2}$ ". The coil L_1 was flat wound with 12 turns of No. 32 copper wire. Its effective diameter was about 1.26 cm. L_2 was 2 cm long, 4.3 cm in diameter, and contained 9 turns of No. 14 copper wire. L_3 contained 9 turns of No. 28 copper wire close wound next to the low end of L_2 .

DRIVING CIRCUIT

The driving circuit which produces the rotating magnetic field in the coils D which in turn spins the rotor is essentially the same as that used by MacHattie.⁶ The 4 coils D are mounted symmetrically around the rotor and opposite pairs are connected in series. The currents in the two pairs of coils differ in phase by approximately 90° and are tuned to the driving frequency by shunting condensers. Figure 3 shows a diagram of the circuit. The output of the transition oscillator is passed through a volume control and then a phase splitting bridge. Each of the phases is amplified and applied to the coils so that the currents in the coils D are 90° out of phase. The frequency used varied from 100,000 cycles per sec. for accelerating the $\frac{1}{2}$ " (3.97 mm) rotor to 500,000 cycles/sec. for accelerating the $\frac{3}{8}$ " (.795 mm) rotor. The constants given in Fig. 3 were those used to accelerate the $\frac{1}{8}$ " (1.59 mm) rotor. It will be observed that the solid steel rotor is actually the armature in an induction motor. In starting the so-called "slip" is very high and gradually decreases as the rotor speeds up. However, the torque at high "slips" gives good acceleration because of the comparative high resistance of the rotor. For example, with the $\frac{1}{8}$ " rotor the acceleration was roughly 30 rev./sec.² with approximately $7\frac{1}{2}$ watts input to the coils D . During some of the experiments the coils D were cooled by a small electric fan.

MEASUREMENT OF ROTOR SPEED

At first some difficulty was experienced in getting a simple accurate method of measuring the rotor speed, but the problem was successfully solved by the apparatus shown schematically in Fig. 4. One-half of the rotor is polished and the other half darkened by dipping in dilute H_2SO_4 , which had been in contact with metallic antimony. This dark layer is very thin and is not destroyed by the highest centrifugal force so

far obtained. Light from an incandescent lamp A is focused on the rotor R and the resulting scattered light focused on the electron multiplier photo-cell P (931A). The variable output from the photo-cell, due to the spin of the rotor, is amplified and fed to one pair of plates of a cathode-ray oscillograph. This frequency of the rotor is then compared in the usual way with the frequency of sweep applied to the other pair of plates of the oscillograph. This sweep frequency was calibrated by a standard frequency oscillator (accuracy from .1 to .01 percent) up to 50,000 cycles per sec. Above this frequency multiples of the calibrated frequencies are used. It is believed that the rotor speeds are precise to three significant figures.

While the above frequency measurements were precise enough for the present experiments, it is highly desirable to have better precision for deceleration measurements. Accordingly, the output of the photo-cell was made to actuate an ordinary scale of 64 counter circuit in such a way that the counter counts the number of r.p.s. directly. This method gives high precision but the present arrangement is not suited for counting above 5×10^4 r.p.s., even with a counter clock, especially wound for the purpose. However, it is hoped to extend its range on up to much higher values.

EXPERIMENTAL RESULTS

As mentioned before the rotors were solid steel spheres (ball bearings). In order to determine the axis of rotation a small flat was ground on each of the rotors. This varied from 1 mil for the $\frac{1}{8}$ "

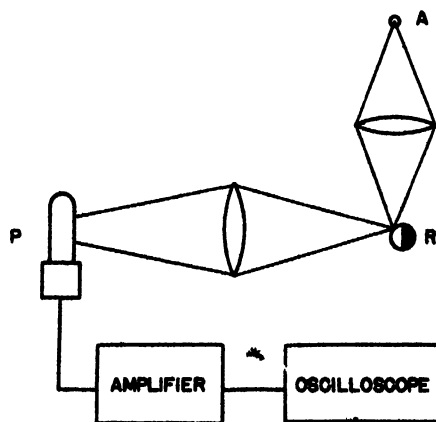


FIG. 4.

TABLE I.

Diam. of rotor mm	Rotor speed r.p.s.	Peripheral speed cm/sec.	Centrifugal acceleration times gravity	Maximum calculated stress lb./in. ²
3.97	77,000	9.60×10^4	4.71×10^7	410,000
2.38	123,500	9.25×10^4	7.20×10^7	385,000
1.59	211,000	1.05×10^5	1.43×10^8	498,000
.795	386,000	9.65×10^4	2.40×10^8	420,000

rotor to 5 mils for the $\frac{1}{8}$ " rotor. This small flat surface became the top of the rotor and hence determined the vertical axis of rotation perpendicular to its surface. The size of the small damping needle *H* (Fig. 1) together with the viscosity of the liquid surrounding it is determined by the size of the rotor. For a $\frac{1}{8}$ " rotor *H* was 8 mils and $\frac{1}{2}$ " long, *N* was 4 mils and $1\frac{1}{2}$ " long and the liquid was "three in one" oil. When the adjustments are properly made no variation in the position of the rotor can be observed. The rotor, the glass vacuum chamber, and the driving coils are surrounded by a wooden barricade which stops the flying parts of the exploding rotor. Although the rotors used were grade *A* commercial steel ball bearings and supposedly free from flaws, a number of them exploded at abnormally low rotor speeds. These rotors showed definite flaws when their parts were examined after an explosion. However, most of the rotors required very high stresses to explode them. Table I gives some of the results obtained just before the rotors exploded.

It will be observed that the maximum peripheral speed for all of the rotors is approximately 10^5 cm/sec. and determines the bursting speed for the steel used. In accordance with expectations the maximum centrifugal field was obtained with the smallest rotor used and for the .795-mm rotor was 240 million times gravity. In the fifth column are tabulated the maximum stresses in the rotors calculated by the method of Chree.⁷ In the calculations it is assumed that the rotors are perfectly elastic. The calculations show that the stresses reach a maximum at the center of the rotor and fall off toward the surface. Actually

⁷ C. Chree, Proc. Roy. Soc. 58, 39 (1895).

the steel is well known to be plastic and it is probably that the material in a small region around the center of the rotor flows and relieves the stresses. Therefore, the calculated stresses are probably well above those actually occurring.

The air pressure surrounding the rotor as measured by an ionization gauge was maintained between 10^{-5} and 2×10^{-5} mm of Hg. The "air" friction on the spherical rotor would produce a deceleration, when the rotor is spinning freely with the driving power off, given approximately by

$$\log_e \frac{N}{N_0} = -\frac{5p}{rd} \left(\frac{M}{2\pi RT} \right)^{\frac{1}{2}} (t - t_0),$$

where N_0 is the number of r.p.s. at time t_0 ; N is the number of r.p.s. at time t ; p is the gas pressure in bars, d is the density of the steel = 7.8 g/cm³, T = absolute temperature; R = gas constant, M = molecular weight of the gas, and r = radius of the spherical rotor. With the 1.59-mm rotor spinning freely at about 120,000 r.p.s. with a pressure of about 10^{-5} mm Hg it required roughly two hours to lose one percent of its speed. Therefore, the order of magnitude of this deceleration can be accounted for by gas friction on the rotor. No doubt the rotors are slightly non-homogeneous magnetically and the supporting field is perhaps not strictly symmetrical. Also, in addition, other small factors such as oscillations in the pick-up coil probably add up to give a small residual drag. However, in practice, since the major portion of the drag apparently is due to gas friction, the deceleration can be made very small. With proper coupling between a rotating magnetic field driven by a piezoelectrically controlled oscillator an almost constant speed rotor should be obtained. Consequently many experiments requiring constant rotor speed as well as high centrifugal fields are made possible.

We are much indebted to Drs. L. G. Hoxton and F. Bader for checking some of our calculations.

A Cathode-Ray Tube for Viewing Continuous Patterns

J. B. JOHNSON

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received June 8, 1946)

A cathode-ray tube is described in which the screen of persistent phosphor is laid on a cylindrical portion of the glass. A stationary magnetic field bends the electron beam on to the screen, while rotation of the tube produces the time axis. When the beam is deflected and modulated, a continuous pattern may be viewed on the screen.

IN cathode-ray oscilloscopes as employed for viewing transient phenomena the luminescent screen with persistent after-glow is often very useful, in that the pattern remains on the screen and can be observed in detail for a short time after the excitation. If the pattern is continuous, while varying in content, then the conventional cathode-ray tubes offer no convenient way of observing the trace without interruption. The requirement of continuity arises, for instance, in

the application to "visible speech patterns"¹ where the time-pitch-intensity pattern of sounds can be viewed for a few seconds as it is continuously produced. The cathode-ray tube to be described here was designed for experimental use in the speech pattern system and other systems with similar requirements.

GENERAL DESCRIPTION

A novel feature of this tube is that the time-axis is produced, not as usual by the electron beam moving across the screen, but by moving the screen horizontally past the beam which remains within a fixed vertical plane, in which plane the deflection takes place. The screen is applied on the cylindrical circumference of the tube, instead of on the end, as shown in the photograph of Fig. 1. The electron gun is contained in the longer neck of the bulb which is coaxial with the screen, and the electron beam is projected along the axis of this neck. By means of a set of coils external to the tube a magnetic field is maintained throughout the large part of the bulb, its direction being perpendicular to the axis of the tube. With suitable strength, this field bends the beam away from the axis within the bulb so that it strikes the screen near its equator. A pair of magnetic deflection coils placed across the narrow neck of the tube permits the beam to be deflected across the screen in a plane through the axis of the tube. Now, with the system of magnetic coils fixed, the tube itself may be rotated about its axis while the beam remains in its fixed plane determined by the stationary bending field. The screen then moves past the phosphorescence-exciting end of

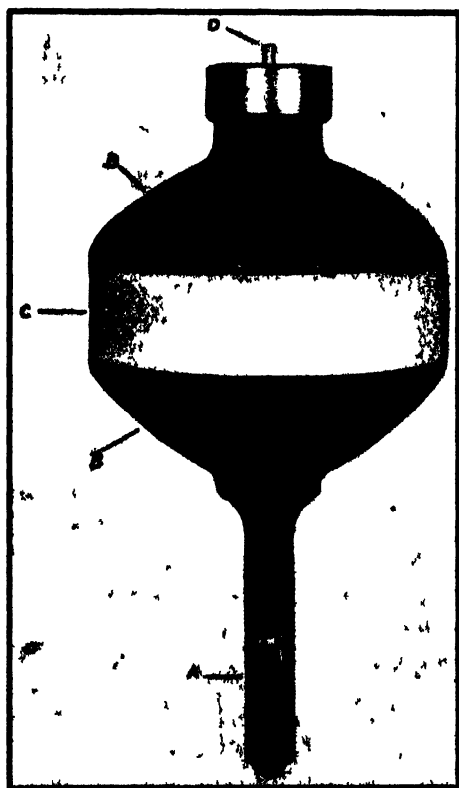


FIG. 1. Photograph of continuous pattern cathode-ray tube. A—electron gun; B—graphite coating; C—luminescent screen; D—high voltage terminal.

¹ R. K. Potter, "Visible patterns of sound," *Science* 102, 463-470, (Nov. 9, 1945). R. R. Riesz and J. Schott, "Visible speech cathode-ray translator," *J. Acous. Soc. Am.* 18, 50 (1946).

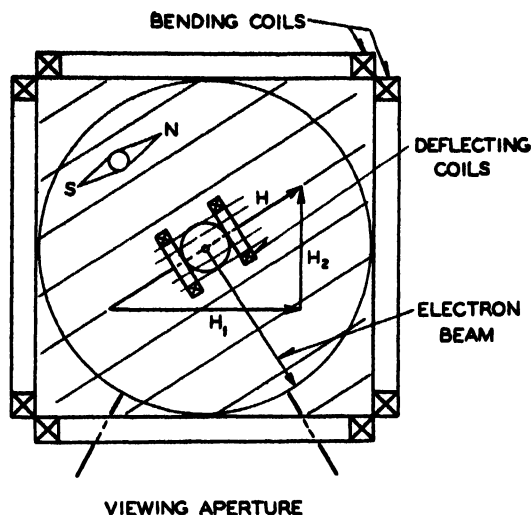


FIG. 2. Horizontal cross section of coil system.

the beam, say from right to left, and a stationary observer sees the history of the trace for the past few moments, the time depending on the rate of rotation. When an exposed area has made a full revolution, the old pattern has become negligibly faint, either by its natural decay-rate alone or with the aid of some wiping-out process such as exposure to light of long wave-length. With these basic features of the tube and its operation made clear, a more detailed description will now be given.

THE BULB AND SCREEN

In order to secure a reasonably long screen the bulb is made large. Its middle portion is a cylinder 10" in diameter and a little over 3" in length. From the cylinder the bulb tapers to a wide short neck at one end and a longer and narrower neck for the electron gun at the other. The latter neck is $1\frac{1}{2}$ " in diameter so as to fit in standard magnetic deflection yokes. The over-all length of the based tube is 21 inches.

The screen is deposited on the inside of the cylindrical part of the bulb, 3" wide. It is of the P7 type of phosphor, with persistent yellow after-glow. The parts of the bulb not occupied by the screen are covered with a layer of colloidal graphite, forming part of the electrode system.

THE ELECTRON GUN

The kind of electron gun is not critical, but electrostatic focusing guns like those of the

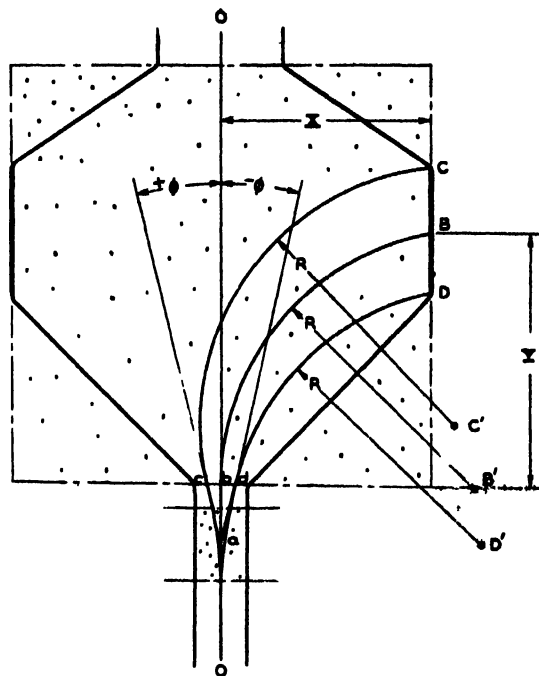


FIG. 3. Vertical cross section of tube in magnetic field. Dots indicate magnetic field directed into the paper. Electron beams OB, OC, OD.

Western Electric '325-A tube or the RCA 7C1P tube have been used, the tube of Fig. 1 having the latter type of gun. The electron beam of this gun can be modulated, the cut-off point being at about -45 volts and the maximum beam current several hundred microamperes. The graphite coating on the bulb forms the final electrode of the gun structure, operated at potentials of 5000-7000 volts. This voltage is brought in through the base at the top of the tube, all other voltages being connected through the smaller base at the gun end. The gun must be lined up very accurately with the axis of rotation of the bulb in order to prevent wandering of the spot while the tube turns.

THE MAGNETIC FIELDS

The arrangement of the coils about the tube is indicated in Fig. 2 which presents a top view of the system. The bending field is produced by a set of four roughly square coils whose sides about equal the diameter of the tube. By using four coils the viewing area can be centered through one of them as shown, with the beam to one side of the observer. A field strength of about 18

oersteds is required for bending the beam to the center of the screen. This field, in a suitable direction, is produced by a mmf of about 640 ampere-turns in each of the side coils and 400 ampere-turns each in the front and back coils, the ratio of field components being $H_1:H_2=8:5$. This field is not very uniform but since the beam makes only a small vertical excursion in it the uniformity is not of great moment.

Below the bending coils and outside their main field is mounted the deflection yoke, shown in the illustration as a pair of air-core coils embracing the neck of the tube above the electron gun. This yoke must be aligned so that its magnetic field is closely parallel to the effective direction of the bending field, for if the fields are not parallel the beam is deflected in an S-shaped pattern instead of a straight vertical line. Distortion is also produced if the deflecting field overlaps appreciably the focusing fields of the electron gun. The deflection sensitivity is of the order of 1 inch at the screen per 50 ampere-turns in the deflection coils.

THE ELECTRON PATHS

In Fig. 3 is shown a cross section of the tube, in the plane of the electron beam, perpendicular to the magnetic fields. For simplicity it is assumed that the magnetic fields are uniform and sharply bounded at the limits of the coil systems, as indicated by dotted areas. The outlines of the tube are shown to scale, with the axis 0-0 and the screen at C-D. Let the radius of the screen be X and the distance from the beginning of the bending field to the center of the screen at B be Y . When the undeflected beam is injected into the bending field it describes the path abB if the field is adjusted so that the radius of curvature of the beam is

$$R = (X^2 + Y^2)/2X.$$

The field is adjusted for this position of the undeflected beam. When now the beam is given a small deflection it enters the bending field as if coming in a straight line from the center point a of the deflecting field, making the angle φ with the axis, and then proceeds in a circle of the same radius R , tangent to the line of entrance, until it strikes the screen. Extreme rays are shown as acC and adD .

On the basis of the idealized field distribution the deflection of the beam from the center of the screen is given very closely by the expression

$$\Delta = Y - y = R \left(\frac{2X}{R} - \frac{X^2}{R^2} \right)^{\frac{1}{2}} - R \left(\frac{2X}{R} \cos \varphi - \frac{X^2}{R^2} + \sin^2 \varphi \right)^{\frac{1}{2}} + R \sin \varphi,$$

where X is the radius of the screen. The middle term in this expression is of the second order in φ so that for small angles of deflection the first two terms nearly cancel and

$$\Delta \doteq R \sin \varphi \doteq R\varphi.$$

Within the range of the screen the deflection is surprisingly linear, in spite of considerable variation in the length of the beam. Astigmatism and defocusing would be expected with the type of path these beams follow, but neither of these has been found very serious.



FIG. 4. Mounting rack for the tube. A—Motor; B—slip-rings; C—deflection yoke; D—driven carriage; E—bending coils; F—screen; G—high voltage terminal.

MOUNTING OF THE TUBE

Two methods have been used for supporting and driving the tube. In the first, the tube is mounted by the bases at each end and driven through the lower base. The preferable method is to support the main body of the tube in a driven cradle, with the bases used only for the electrical connections. In either case all connections are made through slip-rings which are driven with the tube. The rate of drive has been in the range of one to twelve revolutions per

minute, or .5 to 6 inches per second at the screen. Figure 4 shows the tube mounted in a frame with bending and deflection coils and with drive of the second kind.

ACKNOWLEDGMENT

The building of a tube of this kind presented many new problems of structure and procedure, in the solution of which Mr. J. B. Little, Mr. D. A. S. Hale, and Mr. C. L. Loudon deserve much credit.

Improvements in the Stability of the FP-54 Electrometer Tube*

J. M. LAFFERTY AND K. H. KINGDON
General Electric Company, Schenectady, New York
(Received April 23, 1946)

It was found that a considerable portion of the instability in the FP-54 is caused by variations in emission from the thoriated tungsten filament. Operation of the filament at a current which neither activates nor deactivates was found to be a good criterion for adjusting the circuits employing these tubes. Large fluctuations were observed in emission immediately after activation of the filament. Greater stability was obtained by increasing the activation time from 8 to 40 minutes. Filament end shields improved the stability, showing that some rapid fluctuations in emission occur at the poorly activated end portions of the filaments. Long-time drifts were not improved by end shields. Tubes with oxide-coated filaments gave greater sensitivity and less grid current than tubes with thoriated tungsten filaments. These tubes, however, had a tendency to drift. A split type of FP-54 was constructed with a common filament and space-charge grid, but with twin control grids and plates. Both oxide-coated and thoriated tungsten filaments were used. Operation of these tubes in a bridge type circuit eliminated long-time drift and decreased the amplitude of rapid fluctuations.

THE FP-54 plotron is a low grid current vacuum tube commonly used as an electrometer for measuring small currents. This tube has been described by Metcalf and Thompson.¹

The sensitivity which can be utilized with this tube and its associated circuit using a sensitive galvanometer is limited by short-period fluctuations and long-period drifts. It is the purpose of this paper to discuss the results of investigations made to determine the cause of these variations and methods of correcting for them.

CAUSES OF INSTABILITY

The causes of instability in the FP-54 tube and its associated circuit can be listed as follows:

1. External varying electric fields.
2. External varying magnetic fields.
3. Irregular leakage over the surface of the tube envelope.
4. Loose connections and poor rheostat contacts.
5. Microphonics.
6. Variations in supply voltage.
7. Variations in filament emission.

Figure 1 shows a diagram of the circuit used in determining instability. With switch S_2 closed, stability was measured by reading the galvanometer deflection every half-minute for a period of one-half hour and plotting a curve of galvanometer deflections against time. The 1.5-volt cell, with R_1 and R_2 , forms a potentiometer arrangement used in conjunction with S_2 for increasing the grid bias 10 millivolts when making sensitivity measurements on the tube.

The following precautions were taken to elimi-

* An abstract of the material presented in this paper was given at the meeting of the American Physical Society in Cambridge on April 27, 1946.

¹G. F. Metcalf and B. J. Thompson, *Phys. Rev.* **36**, 1489 (1930).

nate or minimize the causes of instability listed above:

1. Tube apparatus, including all wiring, batteries, galvanometer, and shunt were operated in shielded compartments connected by shielded cables.

2. It was not found necessary to shield the tube and circuit from magnetic fields.

3. The envelopes of all tubes tested were washed with alcohol and painted with Aquadag to within $\frac{1}{8}$ inch of the control grid connection. This Aquadag conductor was then grounded at a potential equal to that of the control grid. The tube was operated in an air-tight grounded brass chamber containing phosphorous pentoxide. A rough vacuum was maintained in the chamber during tests.

4. Large heavy duty wire wound resistors, rheostats, and potentiometers were used throughout the circuit. All connections were soldered and made mechanically rigid.

5. The tube was mounted on sponge rubber, and electrical connections to it were made through flexible leads.

6. Three heavy duty Philco storage batteries were used. They are of the high gravity type.

7. Effects due to variations in supply voltage and filament emission will be discussed in detail below.

After the precautions noted above have been taken, the most important remaining causes of instability are fluctuations in storage battery voltage and fluctuations due to random discontinuities in filament emission. Figure 1 is one of the modified circuits of DuBridge and Brown.² The resistances in this circuit may be adjusted in such a manner that the galvanometer reading will be independent of small changes in battery voltage. Further, by balancing the space-charge grid current against the plate current, the galvanometer will remain unaffected by fluctuations in filament emission which change the space-charge grid and plate currents in the *same ratio*.

If for the circuit of Fig. 1, the galvanometer deflection is plotted against the battery current,

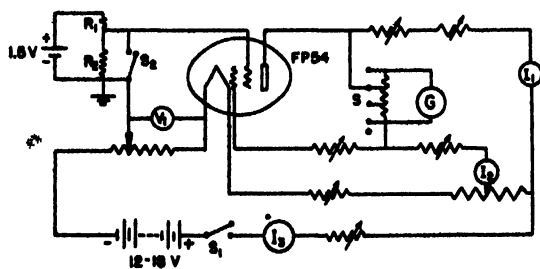


FIG. 1. Diagram of circuit used for testing FP-54 pilotrons.

² L. A. DuBridge and H. Brown, *Rev. Sci. Inst.* 4, 532 (1933). D. B. Penick, *Rev. Sci. Inst.* 6, 115 (1935).

there will, in general, be found two balance points, marked *X* and *Y* in Fig. 2, where $\partial d / \partial I_1 = 0$. If the circuit is operated at point *X*, one would expect any changes in battery current to cause a decrease in the galvanometer reading. However, cases were observed where the galvanometer reading actually increased. Cases were also observed where the galvanometer reading decreased when the circuit was operated at point *Y*. These galvanometer deflections might

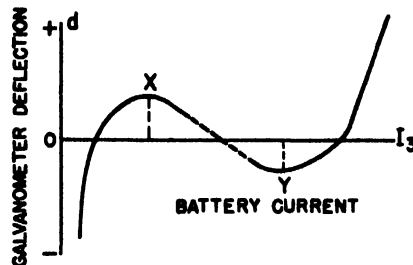


FIG. 2. By proper adjustment of the resistances in Fig. 1, balance points *X* and *Y* may be found where the galvanometer deflection is independent of small changes in battery voltage.

have been caused by increases or decreases in filament emission which did not change the space-charge grid and plate currents in the same ratio, but could not have been caused by battery current changes. It was found that for the batteries used, the galvanometer fluctuations and drift were no worse when the circuit was operated off a balance point. A much better criterion for adjusting the circuit was to operate the filament at a current which neither activates nor deactivates it. This is, in general, between 80 and 95 milliamperes.

ACTIVATION EFFECTS

Thoriated tungsten filaments are usually given the following treatment to make them good electron emitters. First, the filament is heated to 2890°K for a short period of time to clean the tungsten surface and reduce some of the thorium oxide in the tungsten. Any thorium appearing on the surface is immediately evaporated at this temperature. This operation is called flashing. Second, the filament is heated to 2155°K for several minutes. During this activation process metallic thorium diffuses to the surface and a portion of it evaporates producing a surface

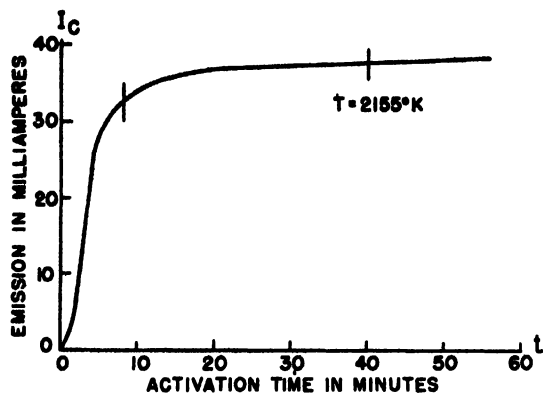


FIG. 3. Filament emission as a function of activation time for the FP-54.

layer of molecular thickness. The thorium layer produced at this temperature gives maximum emission. Third, the filament is reduced to 1800°K for normal operation. However, in the case of the FP-54 the filament is operated at 1650°K to reduce photoelectric grid emission.

For filaments which were just activated, very large and rapid galvanometer fluctuations (100 to 150 millimeters with approximate periods of $\frac{1}{2}$ second) were observed, indicating spasmodic changes in emission. These effects decreased with

time, and after the tubes had been in operation for 24 to 48 hours, they behaved in a more normal fashion.

Figure 3 shows how the filament emission increases with the activation time. Increasing the activation time beyond eight minutes does not increase the emission appreciably. However, it was found that increasing the activation time to 40 minutes gave a marked increase in stability. This is shown by the curves in Fig. 4. A stability curve was first run on a tube which had been activated for the usual eight minutes. The filament of this tube was then flashed at 250 milliamperes for one minute and activated at 150 milliamperes for 40 minutes. A stability curve was then run again after the tube had been in operation for 24 hours. It can be seen that the stability is improved and the sensitivity increased slightly.

Tests were also made on tubes with shields over the ends of the filament (described below), and it was found that the stability improved, if the filaments were activated for the longer period of time.

These facts would indicate that the longer activation time not only activates the cooler end

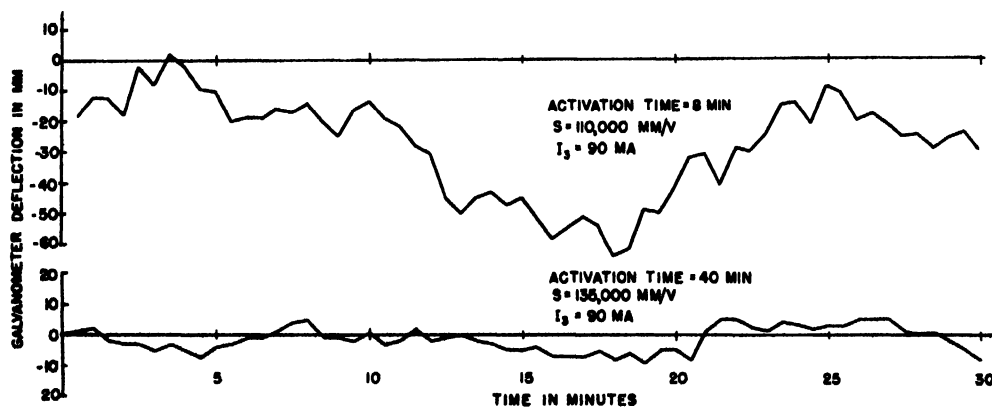


FIG. 4. Effects of increasing the activation time on the stability of the FP-54.

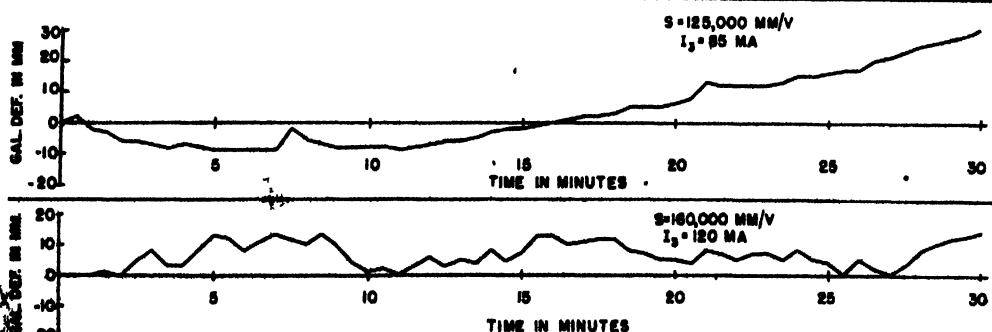


FIG. 5. Stability curve for an FP-54 with filament end shields.

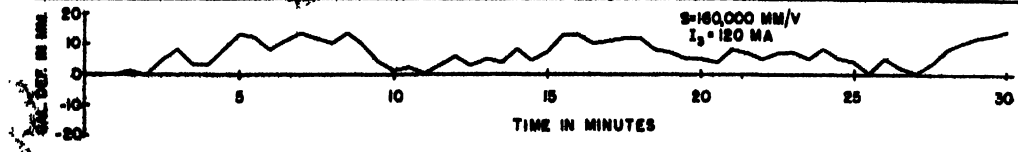


FIG. 6. Stability curve for an FP-54 with an oxide-coated filament.

portions of the filament more completely, but also establishes a more stable equilibrium between the thorium on and below the surface throughout the entire length of the filament.

FILAMENT END EFFECTS

If the instability of the FP-54 is caused by spasmodic changes in filament emission, one would reasonably assume that the effects would be more pronounced near the ends of the filament, where it operates temperature-limited, than along the center portion, where it operates space-charge limited.

Some tubes were constructed, having shields around the ends of the filaments, so that only emission from the hot middle section reached the grids and plate. The shields were connected to the negative side of the filament lead. Figure 5 shows a typical stability curve for this type of tube. Some of the small fluctuations are smoothed out, but the tube still has a tendency to drift over a long period of time. The sensitivity is reduced slightly because part of the active portion of the filament is shielded.

STABILITY OF OXIDE-COATED FILAMENTS

Some FP-54 tubes were constructed, using oxide-coated filaments in place of the thoriated tungsten. The filaments used were of the directly heated type, consisting of a coated one-mil nickel alloy wire 1.5 inches long. These filaments had a rating of 1.5 volts at 120 milliamperes.

A much higher emission was obtained with the oxide-coated filaments operating at a dull red heat (700°C) than with the thoriated tungsten filaments operating at 1374°C. The space-charge-limited emission was also higher for a given anode voltage, because there was less voltage drop across the oxide-coated filament. The increased emission gave the tube a higher mutual conductance and hence a higher sensitivity.

Figure 6 shows a typical stability curve for an FP-54 with an oxide-coated filament. In general, these tubes have a tendency to drift. It was found that if any of the electrode voltages were increased or decreased slightly, the emission would increase or decrease slowly for two or three hours after the change had been made. This effect is very undesirable, since it causes the galvanometer zero to drift.

The grid current for these tubes was about

5×10^{-10} ampere. This is about one-tenth the grid current for the standard tube with the thoriated tungsten filament. This reduction in grid current is undoubtedly caused by the reduction in photoelectrons emitted by the control grid under the action of light from the filament.

SPLIT ELECTROMETER TUBE

The use of two electrometer tubes in a Wheatstone bridge arrangement as a means of compensating for fluctuations in anode voltage and filament current has been described in the literature.³ Because of the symmetrical arrangement, the effect of external disturbances on the grid circuits can also be compensated for.⁴ Random fluctuations in emission from the individual filaments are uncompensated, however.

DuBridge and Brown² have attempted to compensate for fluctuations and drift in filament emission by balancing the plate current against the space-charge grid current. Once this bridge arrangement is balanced, the galvanometer current will remain zero for cases where drift and fluctuations in emission change the plate and space-charge grid currents in the same ratio. This circuit, however, is handicapped by the asymmetry of the space-charge grid and plate. The space-charge grid draws about four or five times as much current as the plate. Under these conditions it was found that drift and fluctuations in plate and space-charge grid current do not change in the same ratio and that emission fluctuations are not entirely compensated for.

Brentano and Ingleby⁵ have suggested that the advantages of the two-tube bridge circuit could be retained, and in addition random filament current fluctuations could be compensated for, if the two tubes were replaced by a single tube with a common filament supplying current to two separate sets of grids and plates. They have described some preliminary experiments showing the feasibility of this method. This principle as applied to the FP-54 structure is described in the experiments below.⁶

³ J. C. M. Brentano, *Nature* 108, 532 (1921). L. A. DuBridge, *Phys. Rev.* 37, 392 (1931).

⁴ J. C. M. Brentano, *Phil. Mag.* 7, 685 (1929).

⁵ J. C. M. Brentano and P. Ingleby, *J. Sci. Inst.* 16, 81 (1939).

⁶ At the time these experiments were made, in the summer of 1940, we did not know of the work of Brentano and Ingleby.

Computations⁷ show that the potential minimum around the filament of the FP-54 is located out from it at a distance of at least 12 diameters. This distance should be great enough to allow the space-charge to redistribute itself almost uniformly around the filament if a change in emission occurred on one side of the filament. Thus, changes in emission current should be nearly the same in all directions. The larger the space-charge density, the more nearly true this

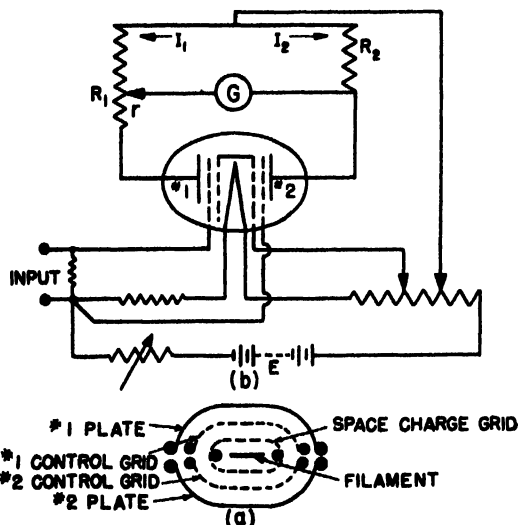


FIG. 7. Basic circuit diagram and electrode structure for the split FP-54.

should be. It was thus decided to modify the FP-54 structure as shown in Fig. 7a, cutting the control grid and plate in two sections. Each section is placed in the arm of a Wheatstone bridge as shown in Fig. 7b.

Assuming for the moment that each section has identical characteristics, a change in filament emission would cause the plate current in each arm of the bridge to change by equal amounts. Thus, if the bridge is balanced, it will remain so independent of changes in filament emission. The input voltage impressed across the No. 1 grid upsets the bridge balance and gives a galvanometer deflection.

For maximum sensitivity, the resistors in the bridge arms should have the same value as the plate resistance of one section of the tube. By using an FP-54 tube split in half, the sensitivity

will be decreased to one-quarter of that obtained for a single tube circuit with the galvanometer in the plate circuit. This is because the mutual conductance is reduced to half-value by splitting the tube, and the bridge arrangement decreases the sensitivity by a factor of one-half.⁸

If the split tube could be built exactly symmetrical, so that each half had identical characteristics, the bridge balance would be independent of variations in the supply voltage E . Actually, such tubes cannot be constructed, and circuits must be used which will compensate for differences in characteristics.

The circuit conditions to be satisfied in Fig. 7 are

$$I_1 = I_2, \quad (1)$$

$$\partial I_1 / \partial E = \partial I_2 / \partial E, \quad (2)$$

and

$$R_1 - r = R_2. \quad (3)$$

Because of symmetry, it was felt that (1) would be the best condition for balancing out drift and fluctuations in filament emission. Condition (2)

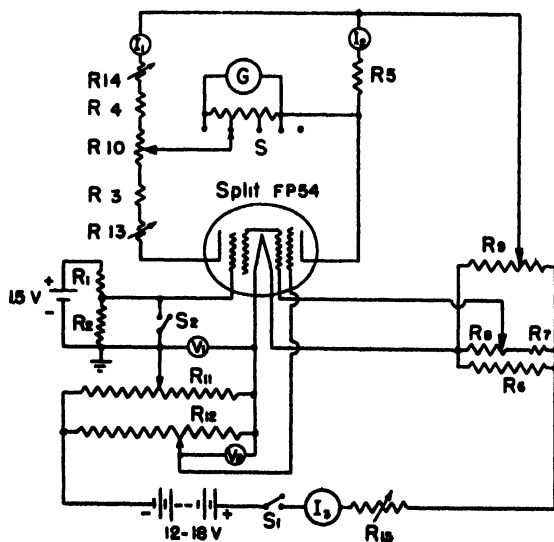


FIG. 8. Split FP-54 circuit diagram.

$R_1 = 150,000$ ohms	$R_{25} = 400$ ohms
$R_2 = 1000$ ohms	$R_{26} = 50,000$ ohms
$R_3 = 0-20,000$ ohms	$R_{27} = 35$ ohms
$R_4 = 0-20,000$ ohms	$S =$ Ayrton shunt
$R_5 = 50,000$ ohms	$S_1 =$ knife switch
$R_6 = 100$ ohms	$S_2 =$ knife switch
$R_7 = 75$ ohms	$V_1 = 0-10$ voltmeter
$R_8 = 125$ ohms	$V_2 = 0-10$ voltmeter
$R_9 = 200$ ohms	$I_1 = 0-100\mu$ ammeter
$R_{10} = 5000$ ohms	$I_2 = 0-100\mu$ ammeter
$R_{11} = 125$ ohms	$I_3 = 0-100$ milliammeter
$R_{12} = 125$ ohms	$G =$ galvanometer (2×10^{-10} amp/mm)

⁷ I. Langmuir, *Phys. Rev.* 21, 419 (1923).

⁸ W. Nottingham, *J. Frank. Inst.* 209, 287 (1930).

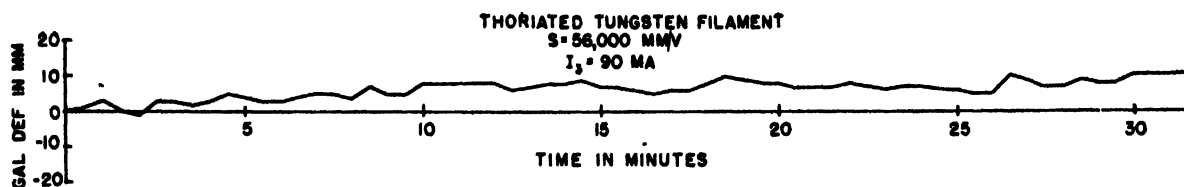


FIG. 9. Stability curve for a split FP-54 with a thoriated tungsten filament.

must be satisfied if the bridge is to remain balanced, independent of small variations in battery voltage. The requirement for bridge balance is given by (3), assuming (1) is satisfied.

Figure 8 shows a diagram of the actual circuit used with the split FP-54 tube. The filament current is adjusted to its rated value by means of R_8 . R_8 and R_9 are used to adjust the space-charge grid and plate voltages, respectively. The grid biases are adjusted by means of R_{11} and R_{12} . The method of balancing this circuit so that conditions (1)–(3) are satisfied is as follows: With the tube operating at normal voltages, the grids are biased equally at -4 volt. The Ayrton shunt is placed on the ∞ tap, and equal resistances are placed in each arm of the bridge. The bias V_1 , on the control grid, is kept constant at -4 volts and V_2 is adjusted until $I_1 = I_2$. R_{13} is decreased until I_3 is increased 10 to 15 percent, and I_1 and I_2 are observed. If I_1 is greater than I_2 , R_{14} is increased; or if it is less, R_{14} is decreased. I_3 is then reduced to its normal value and V_2 is again adjusted until I_1 equals I_2 .

These adjustments are repeated until I_1 and I_2 track, i.e., remain equal for a 10 to 15 percent change in I_3 .

The galvanometer is now connected in the circuit. It may be necessary to insert a resistance R_3 or R_4 , to obtain a balance within the range of R_{10} , but the total resistance of the left arm must be kept constant. With the circuit balanced, the galvanometer is observed for a time to determine if there is any drift. If drift is present in a direction that would indicate I_1 to be increasing more rapidly than I_2 , then R_{13} or R_{14} must be increased slightly. R_{13} gives a fine adjustment of R_{14} . If I_1 appears to be decreasing more rapidly than I_2 , then R_{13} or R_{14} must be decreased slightly. By the proper adjustment of R_{13} or R_{14} , galvanometer drift can be entirely eliminated.

Figure 9 shows a typical stability curve for the split FP-54 with a thoriated tungsten filament. A picture of the tube is shown in Fig. 10. It is constructed from standard FP-54 parts. The control grid and plate have the same dimensions as those of the standard tube, the only difference being that they are cut in two sections. One section of the control grid is supported on the quartz insulators, and the lead is brought out through the top of the envelope. No special precautions were taken to prevent leakage current to the other section of the control grid. An octal base was used for this tube. The standard FP-54 uses a four pin base.

Encouraged by the high sensitivity and low grid current obtained with the oxide-coated filament in the single FP-54 structure, it was decided to try this filament in the split tube. Figure 11 shows a typical stability curve obtained. The sensitivity has been increased to 70,000 millimeters per volt. The objectional drift mentioned above for the oxide-coated filament was entirely compensated for in the split tube. Table I gives the operating conditions

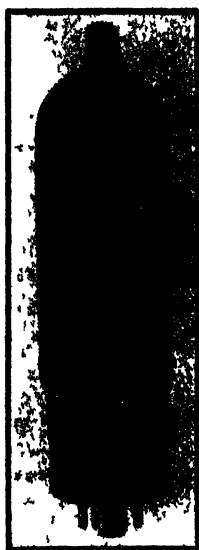


FIG. 10. The split FP-54.

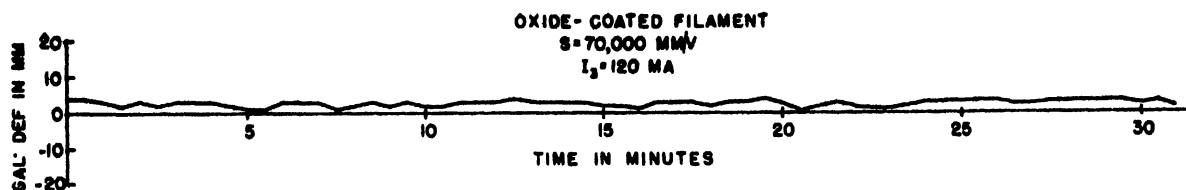


FIG. 11. Stability curve for a split FP-54 with an oxide-coated cathode.

TABLE I. Split FP-54 characteristics.

Filament voltage	1.5 volts
Filament current	0.12 amp.
Space-charge grid voltage	+4.0 volts
Control grid voltage	-4.0 volts
Plate voltage	+6 volts
Mutual conductance	25 μ a/volt
Plate resistance (per plate)	40,000 ohms
Plate current (per plate)	60 μ a
Control grid current	10^{-10} amp.
Control grid capacity	6.5 μ mf

and tube characteristics for the split tube with an oxide-coated cathode.

This tube performed very well indeed, and its stability can be depended upon day after day. Dr. Apker of our laboratory has been using one of these tubes in connection with his photoelectric investigations. His circuit has a sensitivity of 50,000 millimeters per volt and uses a grid resistor of 10^{10} ohms. He reports that the long-time drift is less than 10 centimeters per week and that the amplitude of the short-time fluctuations is about three times greater than the thermal noise of the grid resistor, computed for a band width of one cycle per second.

Calendar of Meetings

November 1946

- 7-9 Conference on Electrical Insulation, Division of Engineering, and Industrial Research, National Research Council, Baltimore, Maryland
- 14-16 Acoustical Society of America, Chicago, Illinois
- 29-30 American Physical Society, Minneapolis, Minnesota

December 1946

- 2-6 American Society of Mechanical Engineers, New York, New York
- 5-7 Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction, Pittsburgh, Pennsylvania (Joint Meeting)

- 17 Institute of Aeronautical Sciences, New York, New York
- 26-28 Geological Society of America, Chicago, Illinois

January 1947

- 3-4 American Physical Society, Los Angeles, California
- 15-17 American Society of Civil Engineers, New York, New York
- 23-25 American Association of Physics Teachers, New York, New York
- 23-25 American Physical Society, New York, New York

February 1947

- 20-22 Optical Society of America, New York, New York
- 24-28 American Society for Testing Materials, Philadelphia, Pennsylvania

A Picturesque Oxidation of Molybdenum Vacuum-Tube Leads

HERMAN A. LIEBHAFSKY

Research Laboratory, General Electric Company, Schenectady, New York

(Received July 10, 1946)

A picturesque formation of MoO_3 on vacuum-tube leads gains in interest and importance because of its close resemblance to a growth of W_6O_{11} on tungsten powder in the electron microscope.

DURING the annealing of certain large vacuum tubes in nitrogen from the factory supply, some oxidation of the six molybdenum leads is always possible. Very rarely this oxidation results in a luxuriant growth of white needles like that in Fig. 1. An almost exact counterpart of this phenomenon on a much reduced scale has been observed by Schmidt¹ on tungsten powder in the electron microscope.

The nitrogen is known to contain oxygen and water vapor, each at a variable pressure usually near several millimeters. The gas enters the lower section of the annealer, which houses the photographed portion of the tube and is separated from the upper by an asbestos partition. The lower section is maintained near 400°C ; in the

upper, where most of the annealing occurs, the temperature is regulated as follows: (1) raised from room temperature to 580° in 45 min. (2) kept at 580° for 30 min. (3) lowered to 400° in 90 min., whereupon the tube is removed. The ceramic spacer (see photograph) keeps the leads aligned during the heating; since this particular spacer had served many tubes, there is no reason to regard it as a source of oxidizing agent.

Qualitative analysis indicated, and x-ray diffraction work proved, that the crystals were MoO_3 . A few simple experiments intended to produce them at will failed altogether; a stunted yellow growth covering the heated surface was all that resulted—apparently Schmidt's work on molybdenum powder in the electron microscope turned out the same way.

Schmidt concluded, by comparing electron-micrographs, that his crystals were W_6O_{11} . The formation, in relatively large amount and in direct contact with the metal, of the highest oxide, but one, of tungsten and of the highest oxide of molybdenum is a striking result—especially in view of the low pressures of the oxidizing agents and of the virtually complete absence of lower oxides. The lower oxides normally form at the surface when a metal is oxidized. It consequently seems necessary to assume that the MoO_3 and the W_6O_{11} are formed above this surface by the action of the oxidizing agent on the lower oxides; oxygen is much more likely to accomplish this than is water vapor.

With neither metal is it possible to predict the occurrence of the pronounced crystal growth at a definite point. It consequently seems plausible that an exceptional condition ("active spot") is required to initiate this growth, which suggests the following picture: At an active spot, oxidation of the metal is rapid enough, to produce an

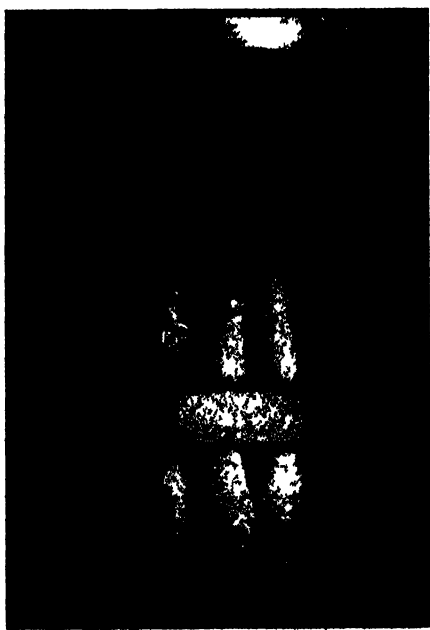


FIG. 1. The oxidation of molybdenum vacuum-tube leads (diameter, $\frac{1}{4}$ inch).

¹ R. W. Schmidt, *Kolloid Zeits.* 102, 15 (1943).

unusually porous structure at the surface. Oxygen continues to diffuse through this structure to generate lower oxides, the oxide superstructure being pushed out into space, where it reacts with oxygen to form the highest oxide (MoO_3 or W_2O_{11}) that is stable under the experimental conditions.

Let us interpret the photograph with this picture in mind. There must be a temperature gradient along the leads during annealing; except near the spacer, the flow of gas is predominantly upward, both because the nitrogen is introduced at the bottom, and because of convection currents. The salient features of the crystal growth with suggested explanations in parentheses are: (1) its lower limit (temperature too low further down) (2) clumps of crystals usually broad above base and narrowest at top (progressively less oxygen as top is approached; note, however, that some crystals have been lost) (3) lower clumps do not quite reach spacer (oxygen "starvation" and physical interference of spacer)

(4) upper clumps begin some distance above spacer (dead space in a gas stream) (5) upper limit (temperature too high and/or oxygen "starvation").

Evaporation and condensation have been ignored thus far, even though MoO_3 and W_2O_{11} are both volatile. Schmidt's electron-micrographs (especially Fig. 1) show thin, long, isolated crystals nearly perpendicular to the metal surface. These crystals grow in a rapidly moving gas stream at a pressure near 10^{-6} mm. The mean free path being large, they could scarcely have grown by condensation. Evaporation and condensation are more likely to have been at work in the *highly branched* growth that Schmidt (Fig. 4) observed on a substratum of W_2O_{11} , although chemical factors are not completely ruled out even here.

I wish to thank Mr. T. G. Crawford, Electronics Department, for calling the crystals to my attention, and Dr. David Harker for supervising the x-ray work that proved their identity.

Luminescence of $(\text{Zn}, \text{Be})_2\text{SiO}_4:\text{Mn}$ and Other Manganese-Activated Phosphors

JAMES H. SCHULMAN*

Sylvania Electric Products, Inc., Salem, Massachusetts

(Received June 10, 1946)

Spectral energy distribution and luminous efficiency measurements of $\text{Zn}_2\text{SiO}_4:\text{Mn}$ and of various samples of $(\text{Zn}, \text{Be})_2\text{SiO}_4:\text{Mn}$ are reported. Quantum efficiencies of these phosphors under 2537Å excitation are calculated from this data. The introduction of beryllium into alpha-willemite causes a decrease of the "green" luminescence band of this phosphor and the growth of a new emission

band with a peak in the neighborhood of 6100Å, the high quantum efficiency being preserved. The general role of manganese as an activator is considered, and the resulting picture used to explain the effect of beryllium and other substituents for Zn^{+2} or Si^{+4} on the spectral properties and on the luminescence decay of manganese-activated materials.

IT is well known that the photo-luminescence color of $(\text{Zn}, \text{Be})_2\text{SiO}_4:\text{Mn}$ phosphors may be varied over a wide range, from green to red, by varying the Zn:Be ratio and the concentration of the manganese activator. Spectral energy distribution data corresponding to these variations are very meager. According to Leverenz and Seitz¹ substitution of zinc by beryllium "flattens the [5250Å-peaked] alpha-willemite

spectrum and shifts it to the red." The measurements of Kröger² on two samples of zinc beryl-

TABLE I. Composition of phosphor samples.

Sample number	ZnO	BeO	Moles MnO	SiO ₂
A	2.0	0.0	0.122	1.18
B	1.8	0.2	0.058	1.11
C	1.8	0.2	0.081	1.11
D	1.8	0.2	0.111	1.17
E	1.8	0.2	0.138	1.20
F	1.6	0.4	0.132	1.19

* Present address: Crystal Section, Naval Research Laboratory, Washington, D. C.

¹H. W. Leverenz and F. Seitz, J. App. Phys. 10, 479 (1939).

²F. A. Kröger, Physica 6, 764 (1939).

TABLE II.

Sample number	Efficiency of luminescence spectrum (lumens per watt)	Quantum ratio	Corrected for Conversion of electrical energy to 2537Å	Unabsorbed 2537; absorption of luminescent light by bulb, phosphor, and lamp ends	Observed efficiency (-5 lumen per watt)	Quantum efficiency
A	517	242	145	124	95	0.77
B	402	174	104	89	79	0.89
C	365	155	93	80	71	0.89
D	345	145	87	75	63	0.84
E	323	135	81	69	58	0.84
F	278	114	68	58	49	0.84

lium silicate at room temperature indicate, on the other hand, that the substitution calls forth a new emission band with a peak at 6100Å. It was, therefore, of interest to study the question further.

The phosphors were prepared by firing mixtures of C.P. ZnO, BeO, MnCO_3 , and $\text{SiO}_2 \cdot x\text{H}_2\text{O}$, in the proportions shown in Table I, at approximately 1200° until their x-ray diffraction patterns showed the same lattice spacings found by Fonda³ in his fusion preparations. This was taken to indicate substantially complete reaction of the components. The spectral energy measurements were made on conventional fluorescent lamps internally coated with the above preparations. Phosphor excitation in such lamps is primarily caused by the 2537Å resonance radiation generated in the low pressure mercury discharge.

The determinations were based on the com-

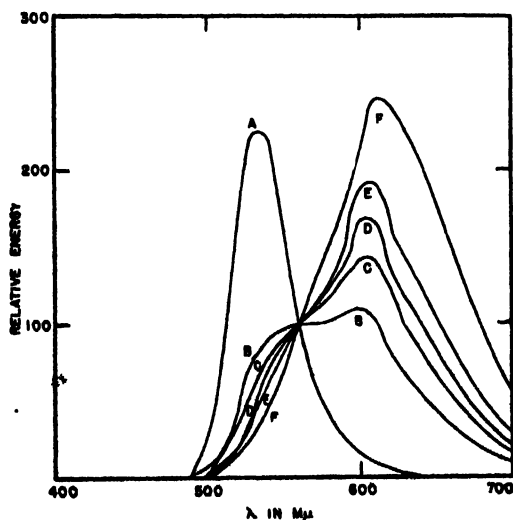


FIG. 1. Spectral energy distribution of luminescence of samples in Table I.

³ G. R. Fonda, J. Phys. Chem. 45, 282 (1941).

parison of the fluorescent lamp output with that of an incandescent source of known spectral distribution. The experimental arrangement consisted of a glass double monochromator equipped with a constant speed motor drive and a revolution counter. By calibration with a high pressure mercury lamp, counter readings were converted to wave-lengths. The receiver was a Cs-Ag-O vacuum photo-tube. Its response was amplified by a Roberts type d.c. amplifier⁴ connected to an Esterline-Angus recording milliammeter. A continuous record of photo-tube response *vs.* counter reading (wave-length) was obtained in this way, the short period of the recording instrument allowing the range from 4000–7000Å to be scanned in seven minutes. The fluorescent lamps were mounted close to the entrance slit without interposition of a focusing lens. To approximate these conditions with the incandescent standard, a block of magnesium carbonate, placed close to the entrance slit and illuminated by the incandescent lamp, served as a secondary source. The standard source was a low voltage, ribbon-filament tungsten lamp operated at a temperature of 2900°K. Slit widths were kept fixed throughout the runs, adjustment of the amplifier gain being employed to obtain recorder deflections of proper magnitude. This adjustment could be made without interrupting the runs. The slit width used corresponded to a $\Delta\lambda$ of 8 mμ at 4000Å increasing to 50 mμ at 7000Å. Results obtained with this arrangement have been found to agree very closely with data obtained from point by point measurements at a constant $\Delta\lambda$ of 10 mμ. Results of the determinations are shown in Fig. 1, E_{6100} being set equal to 100 for all curves.

⁴ S. Roberts, Rev. Sci. Inst. 10, 181 (1939).

Approximate quantum efficiencies of the various phosphors are given in Table II. In the calculation of these values it is assumed that excitation is entirely caused by 2537Å radiation, and that the efficiency of conversion of electrical energy to this radiation is 60 percent in a 40-watt, 1.5-inch diameter, fluorescent lamp. Measurement of the luminous efficiencies of such lamps containing the zinc orthosilicate and zinc beryllium silicate phosphors were made using an integrating sphere, the luminous efficiency of the mercury discharge, about 5 lumens per watt, being subtracted from the measured values to get the efficiencies of the phosphors themselves. Calculations of the quantum efficiencies were made by the method of Thayer,⁵ using his data on the absorption of 2537Å and luminescent radiation by the phosphor coating, visible light absorption by the lamp bulb, and light loss by absorption in the lamp ends. It should be noted that milling of the phosphors is necessary in preparing suspensions for coating the lamps. Since the grinding of phosphors destroys some of their luminescent centers, the efficiencies of the above preparations were originally higher than the values given.

It may be seen from the figure that the introduction of 10 mole percent of beryllium into zinc orthosilicate suppresses the characteristic willemite luminescence and creates a new emission band with a peak at about 6100Å. The quenching of the 5250Å band and the growth of the 6100Å band are both fostered by increase of the manganese concentration at this Zn:Be ratio, and by further increase of the beryllium to 20

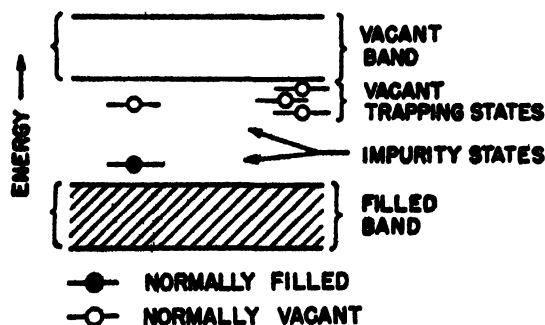


FIG. 2. Energy level diagram for silicate phosphors according to R. P. Johnson.

⁵ R. N. Thayer, *Electrochem. Soc. Preprint* 87, 25 (1945).

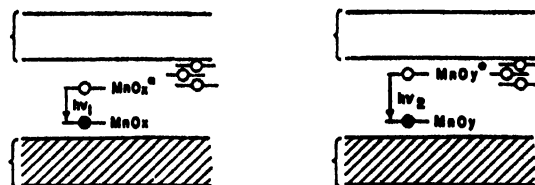


FIG. 3.

mole percent. A high quantum efficiency is maintained through all these changes.

DISCUSSION

The widespread occurrence of manganese as an activator and its association with green, yellow, or red emission bands in a wide variety of host structures suggest that the transition responsible for luminescence takes place within the manganese atom or ion itself or within a coordination group consisting of the manganese ion and its immediate neighboring cluster of anions.^{2,6} This viewpoint is likewise supported by studies of the luminescence decay and of the photo-conductivity of manganese-activated phosphors.^{7,8} From such studies Randall and Wilkins⁸ have concluded that the initial exponential decay of all these phosphors is caused by capture of an electron in an excited state of the Mn^{++} -ion, and is characterized by the lifetime of this excited state. This is essentially the scheme proposed in Johnson's⁷ model of silicate* phosphors, illustrated in Fig. 2, if the normally-empty upper impurity level of this diagram is interpreted as the excited Mn^{++} state. The principal features of Johnson's model are the assumptions that each activator creates *two* localized states in the forbidden region of the host crystal, the lower level being normally filled and the upper level normally vacant, and that the latter is accessible to an electron only if the associated lower level is also vacant at the same time. The transition from the upper to the lower of these two impurity states, resulting in fluorescence, is assumed to be

* S. H. Linwood and W. A. Weyl, *J. Opt. Soc. Am.* 32, 443 (1942).

⁷ R. P. Johnson, *J. Opt. Soc. Am.* 29, 283 and 387 (1939).

⁸ J. T. Randall and M. F. Wilkins, *Proc. Roy. Soc.* 184, 347 (1945).

* Since the published data forming the basis of Johnson's picture deal entirely with manganese-activated silicate phosphors, his model is here considered to represent manganese-activated phosphors in general rather than silicate phosphors in general.

"radioactive in character, little affected by external parameters, and requiring an average time of the order of 10 milliseconds." The model leads to the observed two-stage decay of manganese activated silicate phosphors, accounting for the initial exponential decay and also, *via* the trapping states, for the longer-lived second stage which resembles the decay of sulfide phosphors.

We may incorporate in Johnson's scheme considerations of the luminescence color by assuming further that the average energy difference between the two localized states associated with the manganese activator is a function principally of the number and kind of anions in the immediate coordination sphere of the manganese (cf. Fig. 3), the influence of the host lattice on this separation being of a much smaller magnitude. The manganese-activated phosphors as a class are accordingly considered to be intermediate in type between the (Zn, Cd)S phosphors, where the emitted frequencies are determined by the position of the activator level with respect to the conduction band of the host structure, and the rare earth, Cr^{+++} , and UO_2^{++} -activated phosphors, whose luminescence emission is very markedly characteristic of the activators themselves.

The influence of the coordination number of manganese on the luminescence colors of Mn-activated materials has been considered by Linwood and Weyl.⁶ From a study of the fluorescent silicate, borate, and phosphate glasses they conclude that the "green" fluorescence arises from a fourfold coordination of Mn^{++} with oxygen, while the "red" fluorescence arises from sixfold coordinated manganese. In the first case the manganese is a "glass network former," i.e., it occupies a position in the glass structure similar to that of silicon; in the latter case it is a "glass network modifier," occupying interstitial positions throughout the glass structure, such as alkali or alkaline earth ions do. Corresponding to the first-named role in the glass structure, the bonding of manganese to oxygen is strong, symmetrical, and not greatly affected by thermal vibrations. In interstitial positions, corresponding to the "network modifier" role, the Mn—O bonding is weaker, less symmetrical, and more easily disturbed by thermal vibrations. The effects of temperature and concentration of activator on

the intensities of the green and red fluorescence bands are in qualitative agreement with this viewpoint.

In the extension of their picture to crystalline phosphors, Linwood and Weyl point out that manganese ions situated at crystal defects would have an environment similar to that found at "network modifying" positions in glasses. They reach the conclusion that in general "the long wave emission of Mn^{++} occurs whenever the ion is brought into positions of higher kinetic energy, either interstitial holes or flaws in the crystal structure."

The application of their coordination scheme to the willemite derivative phosphors was attempted by the above authors with partial success. The green luminescence of $\text{Zn}_2\text{SiO}_4:\text{Mn}$ itself was readily and explicitly accounted for by the direct substitution of Mn^{++} for Zn^{++} . Since the Zn^{++} ions in willemite* are surrounded by four oxygens,⁹ this substitution gives rise to the MnO_4 groups responsible for the green emission. The appearance of the red component of the luminescence in $(\text{Zn}, \text{Be})_2\text{SiO}_4:\text{Mn}$ was, however, explained in only general terms, being presumably caused by distortion of the willemite structure by replacement of Zn^{++} ($r=0.83\text{\AA}$) with the smaller Be^{++} ($r=0.34\text{\AA}$), or to changes in the symmetry of the Mn^{++} environment or to lattice defects produced by this substitution. A more specific picture of the possible origin of the "red" centers in the zinc beryllium silicate phosphor is offered in the following discussion.

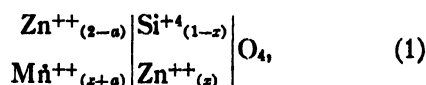
Over rather wide limits of manganese concentration, the spectral energy distribution of the $\text{Zn}_2\text{SiO}_4:\text{Mn}$ luminescence changes but little from the distribution shown in Fig. 1, curve A. The optimum concentration of manganese in this phosphor is about one percent by weight (0.04 mole Mn per mole Zn_2SiO_4). Further increase in activator concentration results in an efficiency loss, together with a broadening of the emission band and a slight shift of its peak toward the red. There is, however, no indication of the growth of a new band at longer wave-lengths.

* The same is true for the Be^{++} ions in phenacite, Be_3SiO_8 , which is iso-structural with willemite. Be_3SiO_8 and Zn_2SiO_4 form limited solid solutions containing up to 30 mole percent Be_3SiO_8 (references 8 and 3).

⁹ W. L. Bragg and W. H. Zachariasen, *Zeits. f. Krist.* 72, 518 (1930).

With a very high manganese concentration, corresponding to complete quenching of the room temperature luminescence, Kröger² was able to observe a bright luminescence at -180° . This low temperature luminescence consisted of both the usual green band and a feebler band peaking at 6100Å, apparently identical with the emission brought out at room temperature by incorporation of beryllium into the structure.*

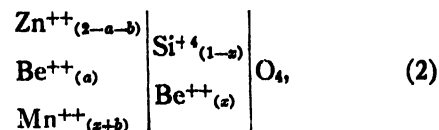
According to the strictest interpretation of the coordination picture, the 6100Å band requires the existence of interstitial manganese with a coordination number greater than four. The appearance of this band in zinc orthosilicate would therefore indicate the presence of such interstitial manganese even in this unsubstituted structure. The preference of manganese for the role of "network former" (fourfold coordination with oxygen) in this structure is shown, however, by the high activator concentration required before the appearance of the 6100Å emission, this concentration being far above the optimum for the green luminescence. The interstitial manganese might be produced by the occasional substitution of Zn^{++} for Si^{++} in the willemite structure, resulting in the formula



where $x\text{Zn}^{++}$ replace $x\text{Si}^{++}$, $x\text{Mn}^{++}$ are in interstitial positions as a consequence, and $(2-a)\text{Zn}^{++}$ and $a\text{Mn}^{++}$ are in the normal fourfold coordinated cation positions. Because of the great difference in size and polarization properties

between Zn^{++} ($r=0.83\text{Å}$) and Si^{++} ($r=0.39\text{Å}$), the extent of this substitution would be small and the green emission band would predominate.*

This type of substitution should be considerably enhanced in the zinc beryllium silicate phosphors, owing to the great similarity in size and polarization properties of Be^{++} ($r=0.34\text{Å}$) and Si^{++} . In this case a possible structural scheme might be the following:



where $x\text{Be}^{++}$ replace $x\text{Si}^{++}$; $a\text{Be}^{++}$, $b\text{Mn}^{++}$, and $(2-a-b)\text{Zn}^{++}$ are in the normal cation positions; and $x\text{Mn}^{++}$ are in interstitial positions demanded by the $\text{Be}^{++} \rightarrow \text{Si}^{++}$ substitution. The Zn^{++} (and Be^{++}) may, of course, compete with the Mn^{++} for these interstitial positions, especially when the Mn^{++} concentration is low. As this is increased, however, the Mn^{++} ions have a correspondingly greater probability of occupying the interstitial holes, and being larger than the Zn^{++} and Be^{++} , may even be preferentially housed in these positions. This would account for the increased 6100Å emission with increase in manganese concentration at a constant Zn:Be ratio. To account for the rapid decrease of the "green" band intensity by introduction of beryllium, it may be necessary to assume also that excitation of the "interstitial- Mn^{++} " centers is more probable than excitation of the "substitutional- Mn^{++} " centers.

Various other possibilities are opened up by considering the effect of different substituents for silicon in the willemite structure. According to the foregoing considerations, replacement of Si^{++} by a suitable tetravalent element, such as germanium, should have no great effect on the spectral energy distribution of the resulting phosphor, for no interstitial manganese ions are made necessary by this substitution. According to the data of Leverenz^{1,10} this is apparently true, the luminescence of $\text{Zn}_2\text{GeO}_4:\text{Mn}$ consisting of a

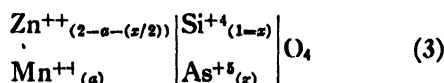
* The red emission made observable by cooling should not be interpreted as the development of a new type of emitting center resulting from structural changes induced by the low temperature. Except for the lattice contraction due to the temperature decrease, we shall assume that the structures of both the red-emitting and green-emitting centers are the same as they would be at room temperature. With the high activator concentrations employed, the thermal vibrations corresponding to room temperature result in a quenching of the luminescence, both red and green, at this temperature. The function of the cooling is merely to reduce the energy exchange between both types of center and the host lattice, in this way making both fluorescence processes possible. It should be mentioned that although cooling and substitution of beryllium for zinc both operate to contract the structure, this cannot be the reason for the change in luminescence spectrum. For if the spectral change were merely a consequence of the lattice contraction, one would expect a gradual shift of the emission peak with decreasing temperature or with increasing beryllium concentration (cf. the (Zn, Cd)S phosphors) instead of the appearance of a new band.

* An identical substitution was postulated by Linwood and Weyl to account for the formation of yellow-fluorescent zinc silicate, a polymorphic form of willemite, with presumably a cristobalite-like structure.

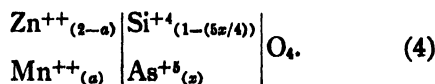
¹⁰ H. W. Leverenz, RCA Rev. 5, 131 (1940).

single band like that of $\text{Zn}_2\text{SiO}_4:\text{Mn}$, with a small shift of the peak to 5370Å.* The same should hold true for any other tetravalent substituent for Si^{4+} , provided that: (a) the willemite structure is retained in the solid solution, (b) the substituent does not act as a "poison" for the luminescence, and (c) the substituent does not introduce a luminescence band of its own.

The substitution for silicon of elements having a higher valence may likewise be approached from this standpoint. For example, Froelich¹¹ has found that incorporation of arsenic or arsenate in willemite causes an "exaggeration" of its phosphorescence. Froelich and Fonda¹² have shown that the action of arsenic consists in increasing the number of trapping states. Substitution as As^{+5} ($r=0.4\text{\AA}$) for Si^{4+} in fourfold coordination with oxygen should be quite possible, giving either



or



The occasional AsO_4 groups distributed through the structure would thus introduce into the crystal positive holes where electrons could be trapped. In (3) a Zn^{++} ion vacancy should occur for every two As^{+5} atoms substituting for Si^{4+} ; in (4) a Si^{4+} vacancy should occur for every four substituting As^{+5} atoms, the positive holes being attracted to the vacant lattice point.

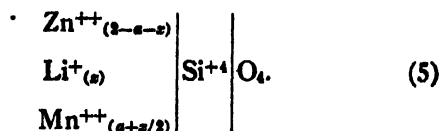
Similar considerations should apply to the substitution of various elements for zinc in the willemite structure. Interstitial cation positions would be demanded by replacement of zinc by a suitable monovalent ion, such as Li^{+1} ($r=0.78\text{\AA}$)* or Na^{+1} ($r=0.98\text{\AA}$), resulting in the following structural scheme.

* Variation of manganese concentration in Zn_2SiO_4 itself causes a shift in luminescence peak from 5230 to 5315Å (reference 1). The shift in peak between Zn_2GeO_4 and Zn_2SiO_4 is thus a minor effect, probably connected with expansion of the lattice by the $\text{Ge}^{4+} \rightarrow \text{Si}^{4+}$ substitution.

¹¹ H. C. Froelich, U. S. Patent 2,206,280 July 2, 1940.

¹² H. C. Froelich and G. R. Fonda, J. Phys. Chem. 46, 878 (1942).

* The fourfold coordination of Li^{+} with oxygen occurs in Li_2WO_4 and Li_2MoO_4 , both of which are "model compounds" of Zn_2SiO_4 (reference 9).



In (5), $x\text{Li}^{+}$ and $a\text{Mn}^{++}$ substitute for $(a+x)\text{Zn}^{++}$ in the normal cation positions, and $x/2\text{Mn}^{++}$ are in interstitial positions. Here again it is possible for the Zn^{++} to compete with Mn^{++} for the interstitial positions, and it may be that the red emission will be observed only at high manganese concentrations. The manganese concentrations required for the appearance of the red band in this case should be lower than that needed in the absence of lithium substitution, however. Trivalent or quadrivalent substituents for Zn^{++} should lead to an increase in phosphorescence by the same method that the substitution of As^{+5} for Si^{4+} enhances this property.

It is hardly necessary to mention that the "red" centers can arise in phosphors, without the rather complicated scheme described above for the willemite, by direct substitution of Mn^{++} for the cation of structures where the latter is normally sixfold coordinated with oxygen. The red luminescence of $\text{Mg}_2\text{SiO}_4:\text{Mn}$ is an example of such a case.

The foregoing considerations may apply also to halide, sulfide, and other types of manganese-activated phosphors, where the appropriate values of n in the MnX_n clusters would have to be determined.

SUMMARY

The effectiveness of manganese as an activator in a large number of different host structures, and the repeated occurrence of either red or green luminescence bands, or both, in such phosphors indicate that the electronic transition giving rise to luminescence takes place within the coordination group of manganese and its immediate anion cluster. This view is supported by the exponential decay characteristics of these phosphors, and by the apparent absence of a connection between the luminescence and the photo-conductivity of these materials. A tentative model of manganese activated phosphors is obtained by combining Johnson's energy level scheme for the decay of silicate phosphors with Linwood and Weyl's results on the effect of

coordination number on the luminescence spectrum of manganese-bearing crystals and glasses. Spectral energy distribution measurements of the luminescence of $\text{Zn}_2\text{SiO}_4\text{:Mn}$ and of various samples of $(\text{Zn, Be})_2\text{SiO}_4\text{:Mn}$ are reported and discussed from the above viewpoint.

ACKNOWLEDGMENT

The writer is greatly indebted to Professor A. R. von Hippel, Laboratory for Insulation Research, M. I. T., for his kind permission to use the monochromator and associated equipment employed in the measurements.

Recent Ozone Investigations

ARTHUR W. EWELL*

Worcester Polytechnic Institute, Worcester, Massachusetts

(Received February 16, 1946)

This paper is supplementary to one which appeared in the December, 1942 issue of this journal. In many applications of bactericidal lamps in air conditioning—such as food preservation, processing, etc.—low concentrations of ozone are desirable to protect surfaces and air pockets not reached by the radiation. A few tenths of one part of ozone per million parts of air is generally sufficient and is obtainable by selecting a lamp with a wall which transmits in addition to the very bactericidal 2537A radiation a few percent of 1850A. The latter is absorbed by the oxygen in the air forming ozone. The reciprocity law (amount of bactericidal action proportional to concentration $\times$ duration of ozone) is applicable for moderate concentration ranges. The bactericidal effect of ozone is nearly independent of humidity for surface infection of foods, etc.,

while high humidity is required for destruction of air borne bacteria. No oxides of nitrogen are produced by high transmission glass or quartz lamps. The ozone output of a lamp is reduced by decomposition by intense 2537 radiation either from the same lamp or from another lamp nearby. Moderate increases in the temperature of a high transmission lamp does not affect the ozone output but increases the dissociation constant until, at 150°F, practically all the ozone is immediately decomposed. The ozone concentration produced by a lamp is substantially reduced by rise in relative humidity, which moderately absorbs 1850 radiation and decomposes much of the ozone produced. Approximate reflection coefficients for 1850 radiation at various surfaces have been obtained by an indirect method.

THIS paper supplements "Production, Concentration and Decomposition of Ozone by Ultraviolet Lamps" which appeared in the December, 1942 issue.¹

OZONE IN AIR CONDITIONING

The chief purpose of the ozone produced by a bactericidal lamp is to supplement the much more germicidal 2537A radiation at infected surfaces not reached by direct, reflected, or scattered radiation, and air-borne infection in stagnant air pockets. Ozone is aided in reaching such pockets by a high diffusion coefficient and in general spreads throughout an enclosure faster than gravitational air currents. On account of the bactericidal contribution of the accompanying 2537 radiation the required ozone concentrations are much lower than required when the ozone is produced by ozonators.

The ozone produced from the oxygen in the air

by radiation of wave-length 1850A, emitted by high transmission ultraviolet lamps, is more effective than the accompanying 2537 radiation in destroying both odor and taste molecules in vapors, and ethylene gas, the latter accelerating ripening of some fruits. The ozone also increases the aroma of small fruits such as strawberries.²

If "house-keeping" is satisfactory, very low concentrations—a few tenths of one part per million (p.p.m.)—are, in regions shaded from 2537 radiation, sufficient to inhibit the growth of surface infection, and air-borne infection—the latter only if the relative humidity is above 40 percent—and to destroy in vapors from foods the molecules which contribute odors and tastes and cause cross-contamination and an objectionable general odor. In the "Tenderay" process for tenderizing meat, 0.2 p.p.m. suffices for control of infection on shaded surfaces; 0.3 p.p.m. is sufficient in cheese aging rooms. Domestic re-

* Consultant, Westinghouse Electric Corporation.
† A. W. Ewell, *J. App. Phys.* 13, 759-767 (1942).

² A. W. Ewell, *Refrig. Data Book* (1946 Edition), pp. 193-199; *Refrig. Eng.* 48, 381-383 (1944).

frigerators require not over 0.1 p.p.m. from special refrigerator Sterilamps.² The 1850 radiation required to produce such concentrations is only a few percent of the total 2537 and 1850 radiation.

Although high concentrations of ozone are objectionable in general air conditioning, low concentrations accompanying 2537 radiation contribute to air purity. Yaglou⁴ found that a concentration of 0.015 p.p.m. in a room crowded with people of "the poorest economic class" reduced body odor to such an extent that to remove the odor the required air changes with outside air could be reduced 50 percent.

The all important factor in applications of ozone is the ozone concentration, c (grams of ozone per cubic meter), determined by the rate of ozone production, per unit volume of inclosure, m (g of ozone produced per minute per m³), and the decomposition coefficient, k ($k = 0.7/T$ where T is the time in minutes, after the ozone is shut off, for the concentration to drop to one-half).

$dc/dt = m - kc$.^{*} At equilibrium $dc/dt = 0$, and $m = kC$, where C is the equilibrium concentration.

COMPARATIVE BACTERICIDAL EFFECT OF OZONE AND 2537 RADIATION

The reciprocity law, which requires that the percent killing be proportional to the concentration $\times$ duration of ozone (p.p.m. $\times$ hr.) holds for limited concentrations. For 50 percent killing of *E coli* on a surface, the writer has found that the product, p.p.m. $\times$ hr., is about 1.8 for ozone concentrations between 0.5 and 3 p.p.m.⁵ Other investigators have found results of the same order of magnitude for mould spores.⁶ The amount of 2537 radiation to produce 50 percent killing of *B coli* is about 5 clicks of a Rentschler Tantalum Click meter.⁷ Therefore, for ozone concentrations of the order of 1 p.p.m., 1.8 (p.p.m. $\times$ hr.) and 5 Tantalum cell clicks have approximately equal bactericidal effects or the bactericidal effect upon surface infection of 1

² A. W. Ewell, *Refrig. Eng.* 50, 523, 524; 560, 561 (1945).

⁴ Yaglou, *Heating, Piping and Air Conditioning* 11, 648-653 (1939).

^{*} Beyond a few inches from the lamp—see Table II.

⁵ A. W. Ewell, *Refrig. Eng.* 41, 331-334 (1941); *Refrig. Eng.* 42, 283-286 (1942); *Refrig. Eng.* 159-161 (1943).

⁶ Smock and Watson, *Refrig. Eng.* 42, 97-101 (1941).

⁷ Rentschler, *J. Am. Inst. Elec. Eng.* 49, 113-115 (1930).

TABLE I.

Relative humidity	Ozone concentration p.p.m.	% kill
28%	4.3	0
36	3.5	0
46	4.	34
52	4.	41
70	3.5	40
92	0.7	83
95	3.5	99

p.p.m. of ozone for 1 hr. is approximately equivalent to 3 Tantalum cell clicks of 2537 radiation.

BACTERICIDAL EFFECT AND HUMIDITY

Rentschler and Nagy⁸ found that the bactericidal effect of ultraviolet radiation is independent of the surrounding humidity for both surface and air-borne infection. There is considerable evidence that ozone is only germicidal in the presence of water.⁹ If bacteria, mould, etc., are on surfaces, which are not dried out, of meat, cheese, fruit, etc., or on a nutrient in a Petri plate, there will always be high humidity for a short distance from the surface. The surface infection is, therefore, associated with sufficient water so that the percentage killing by ozone is quite independent of the relative humidity of the air.

AIR-BORNE INFECTION

Wm. J. Elford and Joan van der Ende¹⁰ found, using very low ozone concentrations and long exposures, that at relative humidities below 45 percent, the germicidal power of ozone upon air-borne infection was negligible, while at high humidities it was very great even for concentrations far below 0.1 p.p.m., if the droplets were not so large as to prevent access to the bacteria, and did not contain excessive absorbing organic matter.

Such dependence upon humidity of the bactericidal effect of ozone upon an air-borne infection is so important in many germicidal applications that it seemed desirable to check these (the only published results known to the writer) by entirely different experimental procedure, with higher ozone concentrations and shorter times.

⁸ Rentschler and Nagy, *J. Bact.* 44, 85-94 (1942).

⁹ Kaess, *Chem. Zeits.* 62, 365 (1938).

¹⁰ W. J. Elford and J. van der Ende, *J. Hygiene, London* 42, 240-265 (1942).

TABLE II.

Distance between axes of lamps	Reduction in ozone
1½ in	23%
1 in	13%
2½	4%
3	0%

An enclosure of about 1.4 m³ was employed. The ozone was produced by a 1000-cycle Siemens & Halske ozonator, the rate of production being controlled by the primary voltage. When a quite constant and uniform concentration was attained, 10–20 cc of a broth culture was sprayed into the top of the booth. The air-borne infected droplets fell slowly. Ten minutes after spraying, sterile culture plates were exposed at the bottom of the box by withdrawing a slide for 1 minute, collecting bacteria which had been sinking through the air and ozone 10–11 minutes. Controls were secured by similar procedure with no ozone. Killing of bacteria during this minute after settling on the plate was determined by exposing plates with known seeding to the same ozone concentration for one minute. The percentage thus killed was subtracted from that of the settling bacteria. The results are presented in Table I. The temperature throughout was close to 70°F.

The manipulation in both this and in the English method was difficult, resulting in some variation in results with both. The agreement is, however, sufficient to justify the conclusion that the *killing of air-borne infection by ozone* is nil at low humidities, moderate at 50 percent relative humidity, and very high when the relative humidity approaches 100 percent. (Note the high killing by only 0.7 p.p.m. at a high humidity.)

DECOMPOSITION OF OZONE BY 2537 RADIATION

Two similar lamps, both producing the same amount of 2537 radiation, but one giving a large ozone output and the other none, were placed side by side in still air. Table II shows the reduction in ozone from the first lamp by additional intense 2537 radiation near the second lamp. With both lamp and silent discharge sources of ozone the output increases when an air current is applied which reduces the length of time that the ozone is subject to the deozone-izing as well as the ozonizing influence near the

source. With brush discharge ozonators the output increases to a constant value at a moderate air speed. With a high transmission lamp as a source of ozone, the yield of ozone increases to a maximum as the transverse air velocity is increased, followed by decrease due to cooling of the lamp. With quartz Sterilamps this maximum occurs at about 100 feet per minute transverse air velocity.

With *increase in humidity* the ozone output from a lamp decreases moderately because of absorption of the 1850 radiation by water vapor. The ozone concentration is further decreased because of increase in the decomposition coefficient.

An *increase in temperature* from 6°C to 31°C (88°F) decreased the ozone produced by lamps to about one-half.¹ Since the temperature in rooms, and particularly ducts, where lamps are installed may be much higher, the following investigations were made:

A high transmission, No. 793, Sterilamp—in an enclosure—was heated in an electric furnace. Tests which were only qualitative indicated a decrease in ozone concentration to zero as the temperature was raised from 70° to 170°F. The current through the lamp only varied between 38.5 and 40 milliamperes. By opening one side of the enclosure the ozone producing 1850 radiation was measured with a Rentschler platinum click meter. Throughout the above range in temperature the 1850 radiation remained constant within 10 percent. (At higher temperatures and mercury vapor pressures, the 1850 output falls off.) Since the ozone output has always been found proportional to the platinum cell readings, it was evident that the decrease in ozone concentrations was caused by increased decomposition. Such was proved to be the case by sending the same initial ozone concentration through an enclosure at different

TABLE III.

Temperature (F)	Ozone concentration
57°	3.
77	2.7
88	2.2
97	1.7
108	1.3
123	0.7
150	0

Above 100° the accuracy was low.

temperatures, then cooling the ozone-air mixture and determining the concentration. (See Table III.)

OXIDES OF NITROGEN

The writer has never found with tetrabase paper any evidence of oxides of nitrogen in the vicinity of lamps producing ozone¹¹ and none has been found in more careful investigations by others. The most comprehensive search for such oxides, using the most sensitive reagents, was conducted by Kohn-Abrest who found that the only gaseous product produced by glass or quartz lamps was ozone.¹²

REFLECTION OF 1850 RADIATION

Reflection coefficients of surfaces for 2537 radiation, at various angles of incidence, can be determined with a Rentschler Tantalum Click Meter. The intensity of reflected 1850 radiation from a bactericidal lamp at moderate angles of incidence is, however, too low for accurate measurements. Knowledge of reflection coefficients for different surfaces is important for choice of reflection materials for fixtures of ozone producing lamps. The following indirect method gave results of qualitative value:

A 9-mm quartz, 4-in. lamp (Fig. 1) was mounted horizontally near the center of a 1.4 m³ enclosure, between horizontal glass plates 6 in. X 8 in. A fan at the lamp level gave an air velocity of 100 F.P.M. The equilibrium ozone concentration was determined by test papers*

TABLE IV.

Black photographic paper	23%
Aquadag coated glass	29%
Bare glass	34%
Alzak	57%

TABLE V.

Distance between plates	Approximate unobstructed volume of air	Average ozone concentration p.p.m.	Ozone concentration compared with no obstruction
No plates	100%	2.0	100%
8 in.	50	1.7	87%
6 in.	40	1.6	80%
4 in.	30	1.4	70%
2 in.	15	1.1	55%

¹¹ A. W. Ewell, News. Ed. Am. Chem. Soc. 19, 1102 (1941).

¹² Kohn and Abrest, Comptes rendus 184, 482-484 (1927).

* It is the writer's experience that the most feasible method for measuring concentrations, 0.4-20 p.p.m., is exposure of potassium iodide-starch papers, of proper

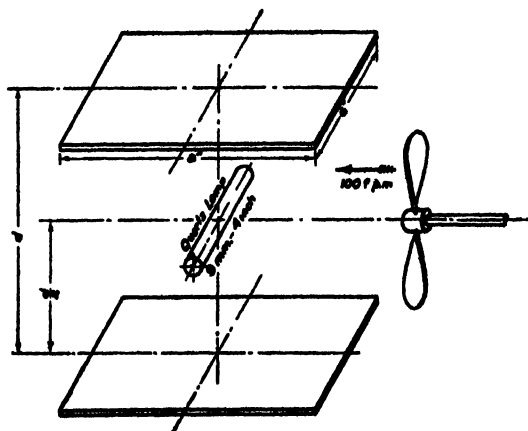


FIG. 1.

suspended from stands at various locations in the enclosure.

The 1850A ozone producing radiation, E , is absorbed by the oxygen of the air (producing ozone) according to the equation¹

$$E/E_0 = e^{-0.044d},$$

where E is d cm farther from the lamp than E_0 . For a distance of 4 in. $1 - E/E_0$ is about 0.4. When the plates were 4 in. above and below (d in Fig. = 8 in.), the unobstructed 1850 radiation was approximately 50 percent. The total ozone producing radiation compared with no plates was, therefore,

$$0.5 + 0.4 \times 0.5 + 0.6 \times 0.5R = 0.7 + 0.3R = C/C_0,$$

when R is the fraction of the incident radiation reflected at angles of incidence, approximately 0-45°, C the ozone concentration in the inclosure and C_0 the concentration without any plates.

Table IV presents the values of R for various surfaces given by this approximate method.

Varying the distance between the alzak plates, which were equal distances from the lamp above and below, gave the results in Table V.

The high values for small unobstructed volumes of air are caused by very great reflection at high angles of incidence.

texture, the ozone concentration being determined from the time, after wetting, required to match one of three standard colors. Papers with colors are calibrated by careful determination of the ozone concentration by titration. Absorption of 2537 radiation as used by Koller [J. App. Phys. 16, 816-820 (1945)] is preferable for high concentrations. Until more sensitive tantalum cells are available, this method is inaccurate for concentrations below 2 p.p.m. For this concentration, at least 200 cm of air and ozone are necessary.

Properties of Welded Contact Germanium Rectifiers

H. Q. NORTH

General Electric Research Laboratory, Schenectady, New York

(Received May 21, 1946)

In the contact rectifier described, the cat whisker has been solidly welded to a germanium surface. The welding process has brought about several extraordinary properties not found in the usual point rectifier: (1) Mechanical stability is achieved without the use of a compound to fill the cartridge. Protection against moisture is gained through the use of a glass-metal seal. (2) The forward dynamic resistance of the microwave unit at 0.5 volt is exceptionally low (2 to 4 ohms), and the back resistance at 0.5 volt is high (0.01 to 1.0 megohm). These values of resistance are accompanied by a low contact capacity at zero bias (below 0.2 μmf). (3) The forward d.c. characteristic follows the law, $I_F = I_0 \exp(\alpha V_F)$ over more than three decades of current. (4) The low forward resistance is responsible in part for conversion *gain* when the crystals are mismatched to microwave radiation at the r-f terminals. (5) Specially welded crystals used as microwave harmonic generators have yielded more than twenty times the harmonic (4-mm wave-length) power output of standard silicon crystals.

1. INTRODUCTION

DURING recent years the need for high sensitivity, low loss detectors at microwave frequencies has reopened the development of contact rectifiers of the point-plane type. The unit described in this paper resulted from an extended investigation carried out in the Research Laboratory of the General Electric Company.¹ The overall purpose of the work was to investigate germanium as a possible improvement over silicon for radar mixers. Only that part of the project which has yielded a new crystal of unusual properties is treated here. An attempt has been made in this presentation to include only enough background of the subject to make the results understandable.

The most important result of our germanium work has been the development of a crystal rectifier in which the cat whisker is firmly welded to a polished germanium surface. The rectifier assembly has been hermetically sealed in a microwave cartridge by means of a glass-metal seal.

The welded microwave rectifier when used as a

mixer at 3-cm wave-length has a noise figure as low as or lower than that of the best silicon mixer crystal. It has in addition some extraordinary properties. In the first place, when an impedance mismatch exists between signal source and crystal with local oscillator applied, *amplification* may actually occur. The increase in signal is accompanied by an increase in noise, however, and the value of the crystal as an amplifier is questionable. Secondly, in the forward direction the unit follows closely the law $I_F = I_0 \exp(\alpha V_F)$ over three decades of current. Here I_F and V_F are the forward current and voltage, respectively. I_0 and α are constants. If the exponent αV_F is changed to $\alpha(V_F - I_F R_s)$, where R_s is a constant (the spreading resistance), the plot of $\log I_F$ vs. $(V_F - I_F R_s)$ is followed closely over more than 5 decades of current. Finally, specially welded germanium crystals have proved superior to commercially obtainable silicon units as generators of both the second and third harmonic of 1.2-cm radiation.

2. THE CRYSTAL CARTRIDGE ASSEMBLY

The cartridge in which all of the microwave tests were made is shown in Fig. 1. Its external dimensions have been made to conform to Army-Navy specifications for the 1N26 type of crystal rectifier. A 0.0015" diameter platinum-10 percent ruthenium wire is first sharpened to a point of less than 0.00002" radius. It is then kinked and spot-welded as shown to the silver-plated steel

¹This work was done in part under Contract No. OEMar-1377 between the General Electric Company and the Office of Scientific Research and Development and under subcontract with the Massachusetts Institute of Technology, Prime contract OEMar-262. The OSRD assumes no responsibility for the statements contained herein.

This paper is a condensation of material presented by the author to the University of California at Los Angeles in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics.

center conductor. The assembly is then sealed to the outer shell with a low loss glass which matches the steel in expansion.

The crystal is a small pellet of treated germanium, soldered to the plug as shown, and polished slowly, first with 600-mesh alundum and finally with magnesium oxide. Electron diffraction photographs indicate that if this slow polishing is done under water, little or no debris remains on the surface.

The germanium used in the welded-contact microwave unit is pure germanium to which 0.2 atomic percent of antimony is added. The addition of this "donor" changes the bulk resistivity from approximately 10 ohm-cm to about 0.005 ohm-cm. The theory of conduction accounting for this 2000-fold change in resistivity has been treated elsewhere^{2,3} and need not be reviewed here. It should be mentioned, however, that

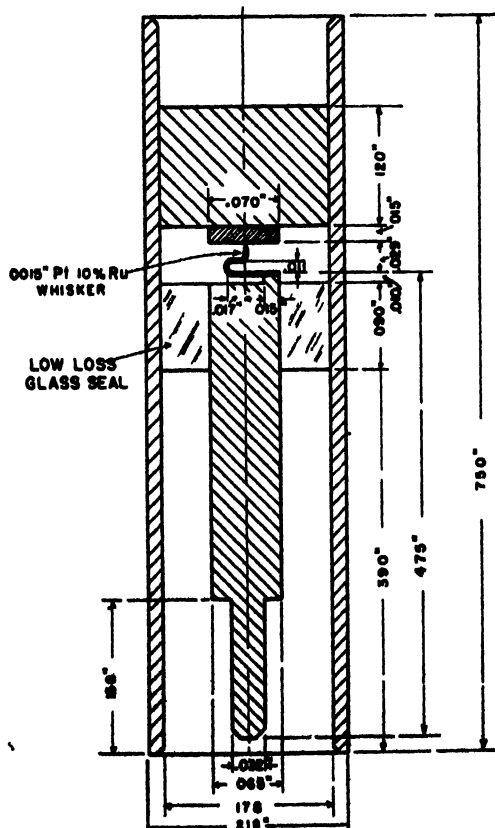


FIG. 1. Germanium crystal cartridge.

² N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Clarendon Press, Oxford, 1940).

³ F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940).

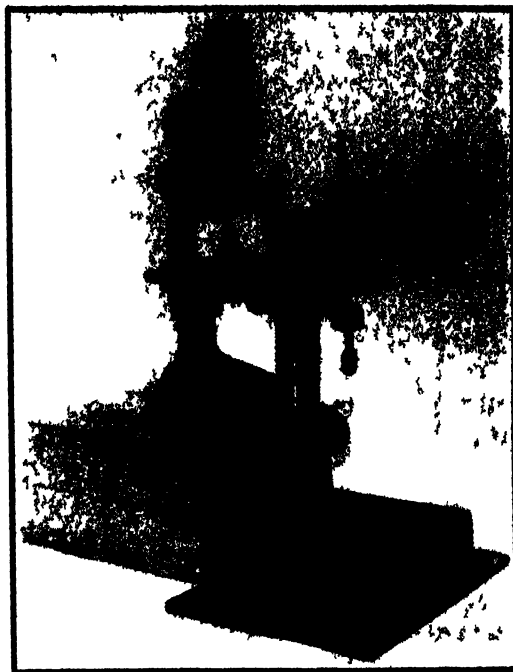


FIG. 2. Assembly to determine force on the contact as a function of whisker deflection.

rigid control of impurities is vital to the construction of good rectifiers. According to theory, approximately 1 atom of antimony in 2000 of germanium is active in producing the 2000-fold change of resistivity mentioned. The remainder of the antimony added is deposited at the grain boundaries.

3. THE WELD AND ITS PHYSICAL PROPERTIES

Early in our research on germanium rectifiers it was found that conversion loss and noise could be greatly decreased by subjecting the units to forward direct currents as high as 100 milliamperes. Contact areas were large at this stage of development, however, and no noticeable welding occurred. Later in the work, condenser discharges were tried, and actual welds were photographed. Because of the large contact areas existing at that time, the condenser-welded units showed poor microwave performance. After the adoption of the present cartridge, a series of careful measurements indicated that contact force should initially be as low as possible and that welding currents should be high for best performance. Best results were obtained when the plug to which the crystal was soldered was pressed in 0.0005" after contact

had been made with the whisker. An optimum welding current of 250 milliamperes was passed through the contact in the forward direction (positive current through the unit from center pin to outer shell) for a few seconds. The duration of the current was not critical. After welding, the unit was sealed at the plug end with low melting solder or glyptal, if the crystal were to be used for other than test purposes.

In order to determine the force on the contact as a function of whisker deflection, or advance of the plug after contact, a standard pin with whisker welded in place was pressed into a hollow cylindrical aluminum rider. The assembly as shown in Fig. 2 was suspended over a polished crystal by a strip attached to a vertical microscope. As the microscope barrel was lowered by either coarse adjustment or vernier, the strip was lowered, and the weight of the pin and rider was transferred to the whisker, now touching the crystal surface. The deflections of the whisker under various weights, added to the rider in successive steps, were observed through the Filar eyepiece of the horizontal microscope shown. Weights were added with the whisker raised from the crystal, and care was taken to avoid ex-

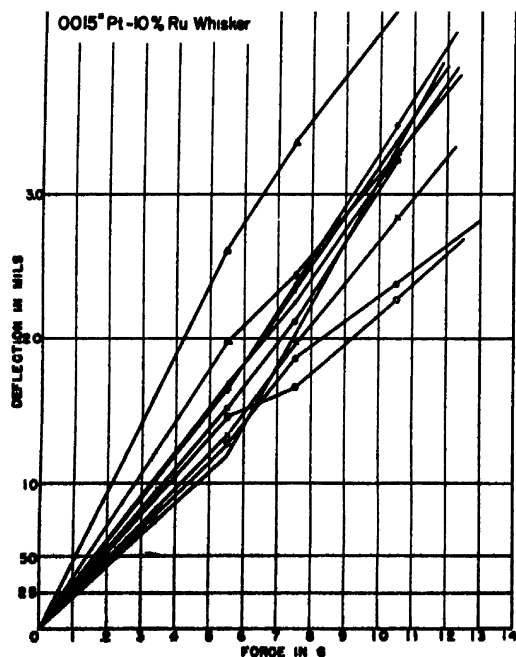


FIG. 3. Curves showing forces required to produce given deflections of standard 0.0015" cat whiskers.



FIG. 4. (a) Sharpened whisker. (b) Whisker which was pulled away from a crystal after welding.

ceeding the elastic limit of the whisker. Figure 3 shows a plot of force at contact *vs.* whisker deflection for several standard 0.0015" whiskers. Breaks in the curve are probably caused by sliding at the contact.

A series of microwave (3-cm wave-length) tests in cartridges had indicated that the best mixer performance occurred when 0.0005" deflections were followed by a welding current of 250 milliamperes in the forward direction for a few seconds. Figure 3 shows that the 0.0005" initial deflection corresponded to a force of about 150 milligrams. After passage of the welding current, an inverse deflection of about 0.0015" or a force of about 500 milligrams was required to break the whisker loose. Figure 4(a) shows a sharpened whisker and 4(b), one which was pulled away from the crystal after welding. The diameter of the whisker below the germanium chip is approximately 0.0002". Thus the welding current density through the contact was approximately 8.0×10^6 amperes per square inch. The tensile stress sustained by the weld was approximately 35,000 lb./sq. in.

The strength of the weld is again illustrated by the stability of a sealed cartridge with welded contact. Such a unit will stand at least a six-foot drop to a linoleum surface without suffering impairment of performance. Many silicon units of equivalent sensitivity will fail under such a test even though filled with the usual paratacopal wax mixture to increase stability and to protect against humidity.

Efforts made elsewhere to weld silicon crystals have, to the author's knowledge, merely resulted in "burn out." This is probably caused by the greater chemical activity of silicon and to its oxidation in air. It is conceivable that silicon

crystals might be welded in an inert atmosphere if chemical combination with the whisker material is avoided.

4. THE D.C. CHARACTERISTIC

Let us consider briefly what d.c. current-voltage characteristic might be expected from a point rectifier. H. A. Bethe has shown that the natural barrier at the contact is thick enough to make diode theory applicable.⁴ His treatment with simplifications can be connected satisfactorily to the phenomena observed in our welded units. Assuming a Maxwell-Boltzmann distribution of velocities for electrons crossing the barrier, he obtained an expression for the electron current density from semi-conductor to metal:

$$j = \frac{1}{2} en v_0 \exp \left[-U/kT \right] (\exp [\epsilon \varphi_a / kT] - 1). \quad (1)$$

Here n is the density of electrons or donor atoms in the semi-conductor and v_0 is the average electron velocity.

$$v_0 = (2kT/\pi m)^{1/2}; \quad (2)$$

ϵ is the electronic charge, k is the Boltzmann constant, and T is the absolute temperature. The factor $\frac{1}{2}$ occurs since only electrons moving toward the barrier get across it. In (1) φ_a is the potential in the forward direction applied to a barrier of height U above the potential in the metal. Bethe has considered the effect of image force upon the d.c. characteristic and has found it important in the case of unwelded contacts. If image force is neglected, $U = \epsilon \varphi_0$ where φ_0 is the contact potential or difference in work functions between the metal and semi-conductor. In this case (1) reduces to

$$j = \frac{1}{2} en v_0 \exp \left[-\epsilon \varphi_0 / kT \right] (\exp [\epsilon \varphi_a / kT] - 1) \quad (3)$$

or

$$I = I_0 (\exp [\epsilon \varphi_a / kT] - 1), \quad (4)$$

where I is the total forward current and I_0 is a constant. In addition to neglecting image force, this equation neglects the probability that electrons tunnel through the barrier. It has also been assumed implicitly that φ_0 is constant over the area of contact.

⁴ This treatment has been described in reports having only limited distribution. A more detailed discussion will be found in a forthcoming book by H. C. Torrey and C. A. Whitmer of the M.I.T. Radiation Laboratory, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1946).

In Eq. (4) the quantity φ_a is not one which can be measured directly. It is the potential applied to the barrier itself, and it does not include the voltage drop due to current passing through the series or spreading resistance of the crystal. The accepted equivalent circuit for a contact rectifier is shown in Fig. 5. Here R_s is the series or spreading resistance, R_x is the non-linear resistance because of the electronic barrier, and C_x is the capacity across this barrier. The spreading resistance is generally obtained by integration over the current path, assuming that the contact consists of a plane circular area between the flattened point and the semi-conductor. It is known that the resistance of such a contact is given by the expression,

$$R_s = \rho / 2d, \quad (5)$$

where ρ is the resistivity of the semi-conductor and d is the diameter of contact.

Since φ_a in Eq. (4) is the potential applied to the barrier of the rectifier, one may re-write the equation as,

$$I_F = I_0 (\exp [\alpha (V_F - I_F R_s)] - 1). \quad (6)$$

Here V_F is the forward voltage applied to the rectifier and α has been written instead of the constant ϵ/kT . It is clear from Eq. (6) that if V_F is large compared to the term $I_F R_s$ in the exponent, and if the exponential is large compared to 1, a plot of $\log I_F$ vs. V_F should be a straight line. Furthermore its slope should be ϵ/kT or about 40 volt⁻¹. In general others have found that α is smaller than this and that it is not constant over a large range of currents. R. G. Sachs, H. J. Yearian, and others at Purdue University have explained these facts by assuming a variation of contact potential over the area of contact.⁴ The welded contact microwave crystals follow an exponential law over a wide range of currents. The value of α obtained is usually more than 60 percent of the theoretical value.

Figure 6 shows a plot of the logarithm of the current against forward applied voltages for a

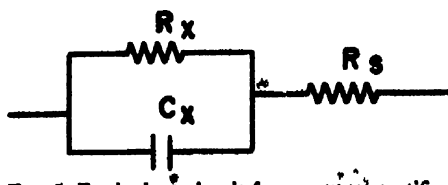


FIG. 5. Equivalent circuit for a contact rectifier.

typical welded contact germanium crystal and for a 1N31 silicon crystal taken at random from stock. The logarithm of the back current is also plotted against the applied back voltage for the two units. The forward characteristic for the germanium crystal is seen to be a straight line over more than three decades of current. The dotted curve is the continuation of the straight portion of the logarithmic characteristic. It actually represents a plot of the logarithm of the current as a function of the voltage applied to the barrier alone as the data points show. To obtain these points, values of R_s were tabulated which would create a voltage drop sufficient at each current to restore the solid curve to the dotted straight line. The average value of R_s was then multiplied by the current at each experimental point, and the voltage drop in the spreading resistance was subtracted from the solid experimental curve. The resulting points fell along the experimental curve as shown, illustrating the fact that the welded contact germanium crystal follows the theoretical expression over five decades of current. For the silicon unit no value of R_s could be found which would restore the experimental curve to a straight line as predicted by the simple theory. We have attributed this fact to the less intimate contact in the silicon unit. The value of α for the germanium unit was 24 volt⁻¹. For this unit I_0 was 9.05×10^{-8} milliamperes. R_s had an average value of 2.59 ohms. Using Eq. (5), the measured resistivity of 0.005 ohm-cm for the germanium, and assuming a contact diameter of 0.0002", one obtains a value of 4.9 ohms. The discrepancy may be caused by the fact that the contact area in a welded unit is probably more nearly a hemisphere than a circular disk. If one makes the reasonable assumption that the current flow near the hemisphere is radial, one obtains the formula

$$R_s' = \rho / \pi d, \quad (7)$$

where R_s' is the spreading resistance of the welded unit, ρ is the resistivity of the germanium, and d is the diameter of the hemisphere. The same value of ρ and d used in Eq. (5) now yield $R_s' = 3.1$ ohms which is in better agreement with the measured value of 2.59 ohms. This low value of spreading resistance accounts in part for the straight portion of the $\log I_F$ vs. V_F plot.

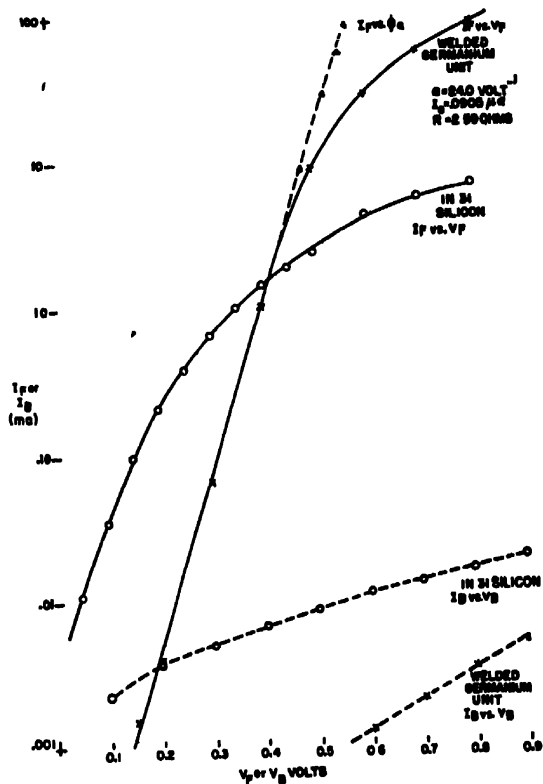


FIG. 6. Comparison of logarithmic characteristics.

Aside from its interest in connection with rectifier theory, the welded contact offers interesting possibilities in applications where resistors are required which follow a single exponential law over a wide range of current. In general the welded units offer such a characteristic over three decades of current.

5. DEFINITIONS OF CONVENTIONAL TERMS FOR MICROWAVE MIXER CRYSTALS

In discussing the microwave performance of mixer crystals three quantities are generally considered: The conversion loss and the noise ratio in turn determine the noise figure which combines the first two quantities. The following definitions are consistent with those proposed by H. T. Friis:⁵

1. *Conversion loss, L_{db} .*—Expressed in decibels, conversion loss of a crystal is defined by the equation,

$$L_{db} = 10 \log_{10} (P_s/P_r), \quad (8)$$

where P_s is the maximum available power from the signal

⁵ H. T. Friis, "Noise figures of radio receivers," Proc. I.R.E., 32, 419-422 (1944).

source at frequency f_s , and P_γ is the maximum available power from the crystal at intermediate frequency γ . P_s is equal to the power absorbed by the crystal and P_γ is that absorbed by the i-f amplifier only when the crystal is matched to the wave guide or coaxial line, and to the i-f load. In general, if the crystal is mismatched to the incoming signal, L_{db} is greater than it is in the matched case since P_γ is reduced.

2. *Noise ratio, N.*—The noise ratio of a crystal is defined by the equation,

$$N = P_s/P_r, \quad (9)$$

where P_s is the maximum noise power available from the crystal to a noiseless resistor, and P_r is the maximum fluctuation noise power available from a resistor at room temperature to a noiseless resistor. Thus if the crystal creates no more noise than a resistor at room temperature, its noise ratio will obviously be unity. The quantity P_r is proportional to the product of the absolute temperature T of the resistor and to the band width, Δf , under consideration. It is given by the equation,⁵

$$P_r = kT\Delta f, \quad (10)$$

where k is the Boltzmann constant.

The noise ratio of a crystal can of course be expressed in decibels as well as the loss. So expressed, it is given by the formula,

$$N_{db} = 10 \log_{10} (P_s/P_r). \quad (11)$$

We shall have occasion to use it in this form.

Expressed in decibels, the noise figure of a crystal is defined by the equation,

$$F_{db} = L_{db} + N_{db}. \quad (12)$$

6. METHODS OF MEASURING LOSS AND NOISE RATIO

The early tests on microwave mixer crystals were made using a set simulating an actual radar receiver. This method of test is known as the "double detection" method. Local oscillator power at frequency ($f_s - \gamma$) was fixed at a level to give 0.5 ma of rectified current through 100 ohms, and the noise power was observed at the output of the receiver. A known signal of frequency f_s was then introduced and increased until the power output of the receiver had doubled. Over-all receiver noise figure was obtained by dividing the known signal power required by $kT\Delta f$. Unfortunately the measurement is a very difficult one when applied to crystals because of the low power levels involved. Furthermore, it does not determine directly the conversion loss and noise ratio of the crystal alone.

To obtain the conversion loss of a crystal directly, a "modulation" method has been used and applied to the production testing of mixer

crystals. R. N. Smith of Purdue University and S. Roberts of the Massachusetts Institute of Technology Radiation Laboratory have shown that modulation on a microwave local oscillator can be used to replace a low level signal in loss measurements. They have shown further that the voltage modulation on the r-f wave can be as high as 10 percent and that a 60-cycle modulation frequency is satisfactory. Loss values calculated from "double detection" results and those obtained by the "modulation" method showed excellent agreement. Roberts has shown that in the "modulation" method, conversion loss is given by the formula

$$L_{db} = 10 \log_{10} \frac{m^2 P_0}{V_L^2/R_L}. \quad (13)$$

Here m is the fractional r-f voltage modulation (60-cycle), P_0 is the average available oscillator power, and V_L is the 60-cycle (r.m.s.) voltage appearing across the load resistance, R_L .

Figure 7 shows a photograph of the microwave "plumbing" used for the loss measurement of the welded units. They were matched to the local oscillator. At the left is a 723-AB oscillator set to a wave-length of 3.2 cm. Following it is a variable resistor strip attenuator and a rotating resistor disk modulator. Fortunately an almost perfect sine-wave modulation is created by the modulator if an eccentric circular disk is used. This greatly simplifies the calculation of the fraction of modulation, m , which can be obtained by measuring maximum and minimum power passed by the modulator. To the right of the modulator in Fig. 7 is a resistor strip attenuator to decouple the modulator from the mixer and a standing wave detector to aid in matching the crystal to the wave guide. Figure 8 shows the d.c. and 60-cycle load circuit used to permit zero d.c. bias voltage regardless of the 60-cycle load. The battery potentiometer was varied until no direct current passed through the load. The load resistor was tapped to permit a variation of 60-cycle resistance as seen by the crystal. The a.c. voltage, V_L , across the load was measured with a Ballantine vacuum tube voltmeter.

The present measurement of noise ratio, although more difficult than that of loss, is simpler than the measurement of noise figure by the "double detection" method. The welded-contact

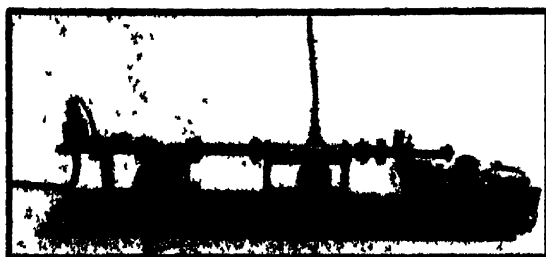


FIG. 7. Photograph of microwave plumbing used for loss measurement.

crystals described here were measured, in the matched case, on a standard noise set designed at the Massachusetts Institute of Technology Radiation Laboratory. The mixer was altered to permit matching each crystal to the padded wave guide. Since this noise set has been described in detail elsewhere,⁶ its principles of operation will be merely outlined here. The set consists of a narrow-band 30-megacycle amplifier whose output meter reading is proportional to the power dissipated in its input resistance. An impedance matching circuit connects the amplifier input to the crystal under test. It is of such design as to permit the testing of crystals having i-f impedances between 150 ohms and 600 ohms. Using a known diode noise source, one may calibrate the output meter directly in terms of the noise ratio of the crystal at the input.

7. PERFORMANCE DATA ON WELDED MIXER CRYSTALS MATCHED TO THE LOCAL OSCILLATOR

Several hundred welded-contact germanium crystals were tested for loss by the modulation method and for noise by the method just described. In the loss measurement 1 milliwatt of local oscillator power was matched into the units. The crystals fed a 400-ohm 60-cycle load resistance. The d.c. bias was set to zero, although with 1 mw of available local oscillator power, bias was not critical. Figure 9 shows the loss distribution of a typical group of 177 welded crystals matched to the local oscillator. Figure 10 shows the noise distribution for the same group excluding 13 crystals which did not give sufficient current. During the noise measurements the units were matched to the local oscillator which was

⁶S. Roberts, 1N23 Noise Measuring Set Type 7438, Radiation Laboratory Report M-190 (Dec. 21, 1944).

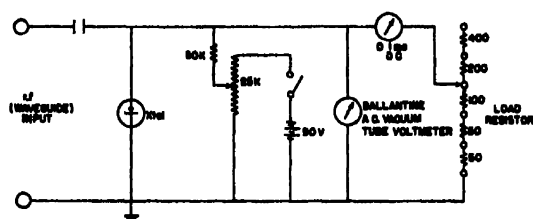


FIG. 8. Load circuit for loss equipment.

attenuated to give 0.5 ma of crystal current in each case. The values of loss and noise ratio for these units welded with 250 ma of d.c. are considerably lower than those obtained earlier for unwelded units. In general the crystals having low conversion loss are those with low noise ratios as well. Some improvement over the values of noise ratio plotted can be obtained by lowering local oscillator level. If bias is used, no appreciable increase in loss accompanies the decrease in noise.

8. CONVERSION GAIN IN WELDED UNITS MISMATCHED TO THE LOCAL OSCILLATOR

During the measurement of conversion loss of welded contact germanium crystals, it was noted that somewhat lower losses could be obtained when the crystals were mismatched to the local oscillator, creating a voltage standing wave ratio of about 5 to 1 in the wave guide ahead of the crystal. It was found that if an i-f load above 1000 ohms were used, the conversion loss in db often became negative or conversion gain occurred. Up to this time it had not been known that a crystal could furnish more power to the i-f load than it received as low level signal in the presence of a local oscillator. The Massachusetts Institute of Technology Radiation Laboratory, and the University of Pennsylvania, as well as the General Electric Company started groups working with the welded units furnished by the author, and the subsequent evaluation of the effect arose through the joint efforts of all concerned.

M. C. Waltz and R. H. Dicke of the Radiation Laboratory found that a resonant circuit attached to the i-f terminals of these units (fed by a mismatched microwave oscillator) would oscillate at the resonant frequency of the circuit, at least up to 34 megacycles. The presence of negative resistance was thus indicated. Figure 11(a) is a

photograph of the characteristic of a crystal with about 1 milliwatt of r-f applied but mismatched to the crystal. There are actually two regions of negative resistance observed, one near zero bias and one at a negative bias of about 0.9 volt.

To explain the conversion gain of the welded units, it is only necessary to explain the mechanism creating negative resistance at the i-f terminals. This can be seen most easily by a consideration of Fig. 11(b). It is a photograph of the characteristic of Fig. 11(a) when 60-cycle modulation (6 percent) was applied to the r-f. The blurred portion in the negative resistance region consists of a series of characteristics with r-f power as the parameter. Consequently a load line drawn through this region defines the limits of the i-f voltage swing for the modulation used. The load line actually drawn in Fig. 11(b) is nearly parallel to the negative resistance portion of the characteristic. Consequently a very small signal or r-f modulation will cause a large voltage swing along the i-f load line. This swing is almost independent of the r-f signal level. Thus, conversion gain can result. (The term gain is actually not uniquely defined when such a characteristic is

obtained.) H. C. Torrey of the Radiation Laboratory has shown that with reasonable assumptions, linear network theory will account for the negative i-f conductance. His argument requires the variation of contact capacity with bias and low spreading resistance, both of which occur in welded germanium rectifiers.

W. C. Hahn of the General Electric Research Laboratory, using the assumption of a non-linear capacity set up a low frequency analogue to simulate the microwave performance of welded contact crystals. With it he was able to duplicate all of the phenomena observed with the microwave unit. The circuit producing the required characteristics is shown in simplified form in Fig. 12. This circuit approximates rather closely the one found in the microwave case. The resonant circuit including L and C exists in the wave guide mixer tuned to cause a large standing wave ahead of the crystal. L_w corresponds to the inductance equivalent to the combination of the glass bead and the cat whisker in the crystal cartridge. R_s corresponds to the spreading resistance of the germanium. R_b and C_b are the barrier

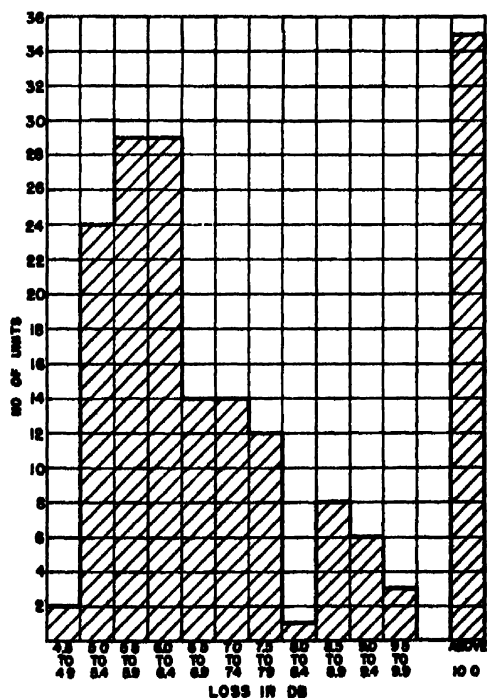


FIG. 9. X-band loss distribution for 177 welded crystals matched to local oscillator.

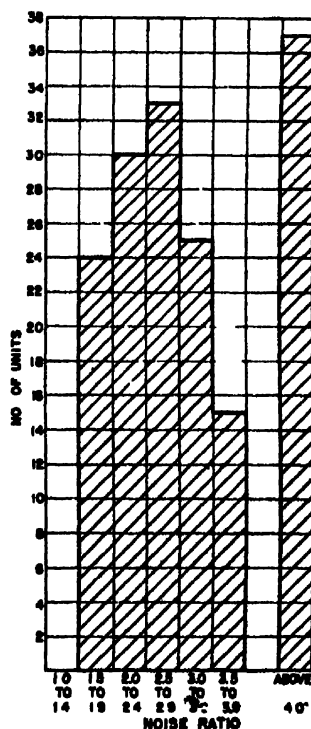


FIG. 10. X-band noise distribution for 164 welded crystals matched to local oscillator.

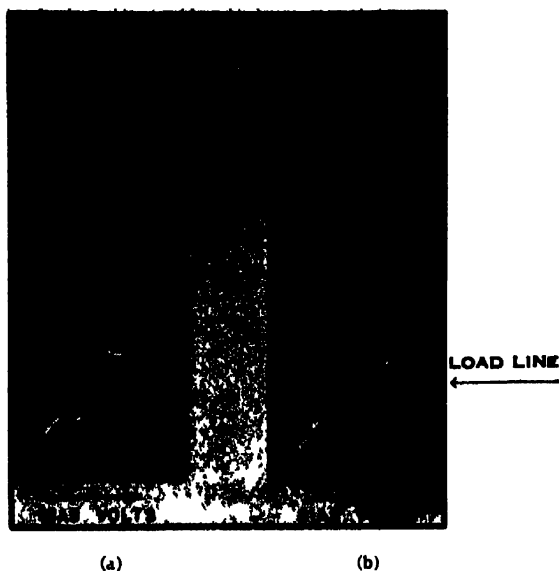


FIG. 11. (a) Characteristic of a crystal with about 1 mw of r-f available but mismatched to crystal. (b) Photograph of the characteristic in (a) when 60-cycle modulation was applied to r-f.

resistance and capacitance varying with bias. C_s was obtained and varied in the low frequency analogue with a reactance tube circuit.

With the circuit of Fig. 12 Mr. Hahn readily obtained negative "i-f" resistances when R_s was sufficiently small and L_w was sufficiently large. The necessity for a low value for R_s agrees with the finding that silicon microwave crystals (having large spreading resistance) do not show negative i-f resistances. A few microwave crystals with short, L-shaped whiskers of decreased inductance were made. Tests confirmed Mr. Hahn's finding that L_w must be high, since no negative resistances were obtained. Rigorous analytical treatment of the circuit of Fig. 12 is very difficult. The general mechanism whereby conversion gain is obtained can be seen from the circuit, however. In the presence of modulation or r-f signal, the changing capacity, C_s , and resistance, R_s , throw the L - C circuit in the wave guide in and out of resonance. This varies the bias voltage at the modulation or intermediate frequency. The result is a transfer of local oscillator energy to the i-f terminals of the crystal.

It should be pointed out here that the negative resistances obtained are only possible because of the low spreading resistance and the circuit constants of the welded contact crystal. They are

not caused by thermal effects as are the negative resistances in crystals heretofore described.⁷ The conversion gain described for welded-contact germanium crystals depends upon the presence of a mismatched local oscillator, part of whose energy appears at the i-f terminals of the crystal when a signal is applied to the r-f terminals.

9. THE PERFORMANCE OF MISMATCHED CRYSTALS AS A FUNCTION OF BIAS

Since the welding process had made possible conversion gain in germanium crystals, a knowledge of crystal noise under conditions of mismatch became essential. For the measurement equipment on hand was used. The modulation set (Fig. 7) and the standard noise set were modified for the purpose. The equivalent circuit in Fig. 13 was assumed for the crystal as seen from its i-f terminals. Here i_n and i_s are constant current noise and signal generators for a given setting of the r-f tuning elements and a given d.c. bias. b represents the i-f susceptance of the crystal and g_i its i-f conductance. The circuit is commonly used in mixer representation and need not be justified here. The method of measurement was not conventional, however, and needs further discussion.

For the loss measurement mixer tuning was fixed on the modulation set to give the desired characteristic during the application of 1 milliwatt of available local oscillator power. The 1000-cycle admittance of the crystal to a low level signal was measured as a function of bias. It is known that if the r-f tuning elements are close to the crystal, its 1000 cycle admittance is equal, for all practical purposes, to its i-f conductance. The conductance, g_i , for a typical crystal has been plotted as a function of d.c. bias in Fig. 14. Note that the conductance becomes negative at a positive bias of 0.25 volt. To determine the value of i_s for a given signal input power, with the

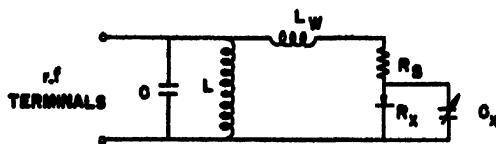


FIG. 12. Low frequency analog of a negative resistance crystal.

⁷ E. C. Cornelius, "Germanium crystal diodes," *Electronics* 19, 118-123 (1946).

mixer attached to the modulation set, the i-f load was set at 100 ohms. The 60-cycle voltage, V_L , across this load was measured as a function of d.c. bias. One milliwatt of local oscillator power was made available. From the value of V_L and g_{if} , i_s was determined for the equivalent circuit.⁸ It may be seen from the circuit that the power *delivered* to the load conductance g_L is given by the expression

$$P_L = \frac{i_s^2}{(g_{if} + g_L)^2} g_L, \quad (14)$$

when susceptance is negligible or tuned out. The usual definition of loss refers to power *available* from the i-f terminals of the crystal. When g_{if} is negative, power "available from the crystal" has no meaning. Consequently loss has been re-defined for mismatched units as the ratio of available signal power to power *developed* in the load conductance, g_L . In Fig. 14, L_{db} is plotted as a function of d.c. bias voltage. For positive values of g_{if} , the usual definition of loss was used assuming $g_L = g_{if}$. A particular load of 0.0002 mho was considered in the negative resistance region since that is the order of grid-cathode conductances at intermediate frequencies.

The measurement of noise ratio was made on the standard noise set, but the input circuit was arranged as shown in Fig. 15. The equivalent crystal circuit of Fig. 13 has been included. The method used was essentially one of substitution. Crystal noise was compared with that of the temperature limited diode, D , shown in the figure. The inductance L_A was used to tune out the capacitance of the diode, that of the 6AC7, and the susceptance of the crystal. The wave guide attenuator was set so that 1 milliwatt of local

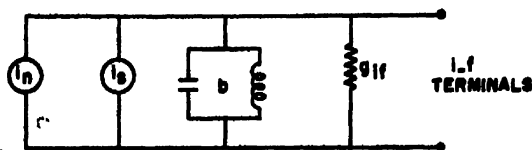


FIG. 13. Equivalent i-f circuit of a crystal rectifier.

⁸ In measuring V_L in the region of low or negative i-f conductance, it is important either that the percent modulation be small enough or that the load line be steep enough to prevent distortion of the wave due to a large voltage swing at the i-f terminals. Six-percent modulation and a load of 100 ohms were found to be satisfactory.

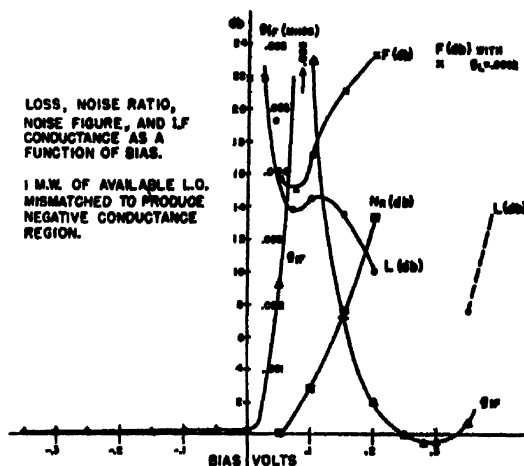


FIG. 14. Typical data on a welded crystal.

oscillator power was available as in the loss measurement. Wave guide tuning was left the same in the two measurements, so that i-f impedances were known for the noise measurement. The noise output for each crystal was compared with the known diode output, leaving the crystal in place when the diode was turned on, to insure the proper value of g_{if} in the measurement. In the negative resistance region, noise ratio was defined as the power *delivered* to the i-f load divided by $h\nu\Delta f$ (cf. Eq. (9)). For positive values of g_{if} the usual definition of noise ratio was used assuming $g_L = g_{if}$. In Fig. 14 N_{db} has been plotted as a function of bias. The 2000-ohm grid resistance of the amplifier prevented oscillation of the input circuit in the presence of the negative crystal resistance.

In Fig. 14, F_{db} , which is the sum of L_{db} and N_{db} , is also plotted. A continuous curve for this quantity has been plotted in the region of positive g_{if} . In the region of negative g_{if} , a typical value for F_{db} has been plotted from typical values of L_{db} and N_{db} in this region. For this crystal as for all others tested with mismatched local oscillator, N_{db} goes up as L_{db} goes down in such a way that F_{db} remains higher than its value obtained when the crystal is matched to the local oscillator.

The conditions under which these data were taken should be carefully specified. L_{db} , N_{db} , and F_{db} were obtained for the case in which signal, local oscillator and image were all mismatched alike. The modulation method was assumed applicable to these crystals. It should be remembered

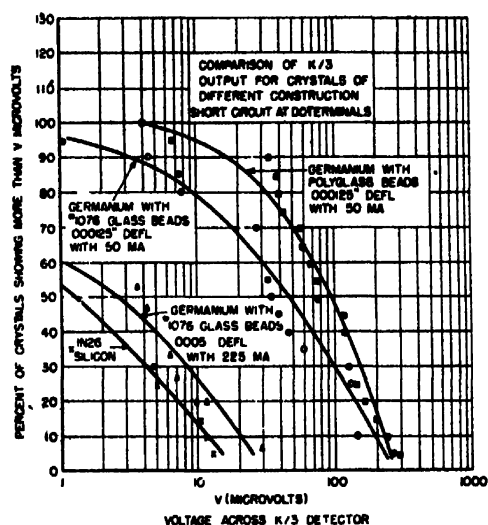


FIG. 17. Comparison of $K/3$ output for crystals of different construction.

1N26 silicon crystals was compared to that of each of three groups of welded germanium units. The first group was constructed and adjusted in approximately the same manner as the welded contact microwave units previously described. The 0.0005" whisker deflection was followed by a welding current of 225 ma. The second group was the same except for the adjustment and welding current. The whisker deflection of this group was 0.000125", and the welding current was 50 milliamperes. The third group of welded contact germanium crystals was like the second in adjustment and welding current, but the usual glass beads were replaced by beads of "Polyglass D," a thermoplastic material having a lower dielectric loss than the glass. The relative harmonic output of the three groups and that of the standard silicon units is shown in Fig. 17. The percentage of crystals showing more than V microvolts at the detector has been plotted against V , the detector voltage. It is apparent from the curves that our standard welded contact microwave crystals are somewhat better generators of 4-mm waves than the silicon units, and that more than twenty times the output of the silicon units can be obtained from welded germanium crystals with special contacts. Some improvement was noted when the "Polyglass"

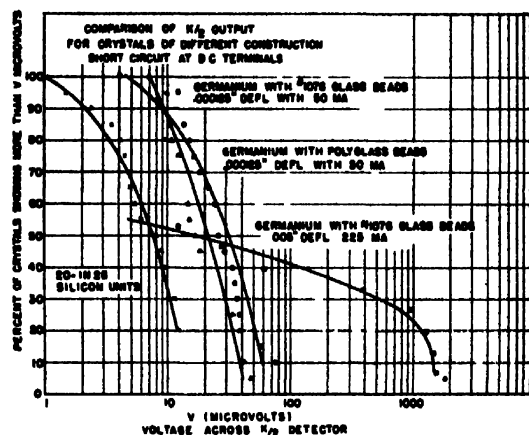


FIG. 18. Comparison of $K/2$ output for crystals of different construction.

beads were used, presumably because of the decreased absorption of the harmonic.

Because of the possibility of obtaining much more second than third harmonic, the same groups of crystals were compared in a holder similar to that for the third harmonic. Figure 18 shows a comparison of the second harmonic output from the groups of crystals previously tested. It is interesting to note that the light-contact germanium units were only slightly better than the silicon units in second harmonic generation. It is also interesting that several of the germanium units welded in the usual manner showed very high second harmonic output.

11. ACKNOWLEDGMENTS

The author wishes here to thank Dr. A. W. Hull of the General Electric Research Laboratory for his encouragement throughout the project and for his suggestions concerning this presentation.

Thanks are extended to Dr. H. C. Torrey of the M.I.T. Radiation Laboratory, who first pointed out the linearity of the logarithmic characteristic of our units, and whose suggestions were extremely helpful during the course of this work.

To Miss M. E. Kershaw gratitude is expressed for her suggestions during the course of the work and for her patience in helping with many tedious operations and tests.

Fine Structure of Viscose Rayon*

H. G. INGERSOLL

Chemical Department, E. I. du Pont de Nemours & Company, Wilmington, Delaware

(Received May 31, 1946)

Quantitative x-ray diffraction methods have been used to study the orientation of the cellulose chains and lateral ordering around the long chain axes in various commercial and experimental rayons. Lateral order can vary independently of orientation in cellulose, as is the case with other linear polymers such as the polyamides. The degree of lateral order obtainable by aqueous annealing of regenerated cellulose appears to be restricted by the structure originally set up. As noted by Sisson, the well-known relations between yarn tenacity and orientation and elongation to break and orientation are not unique. At a given orientation, a range of tenacities and elongations are

possible, depending upon the influence of other variables. Study of a series of yarns prepared under constant mechanical conditions showed that elongation to break increased with lateral disordering of the cellulose chains at constant orientation. The data support Baker's proposal that local chain disordering in any linear polymer favors higher extensibility. Some yarns of essentially the same orientation and lateral order have shown different properties indicating that other as yet undefined structural variables also have an important effect on physical properties.

INTRODUCTION

EXTENSIVE investigations over a period of years by Herzog, Hengstenberg, Meyer, Mark, Sisson, and many others have led to acceptance of the hypothesis that cellulose consists of both organized or crystalline and unorganized or amorphous regions.¹ The arrangement and dimensions, relative amount or perfection of the organized regions and the nature and arrangement of disordered regions vary widely depending upon the source and history of the sample. This heterogeneous order-disorder structure of cellulose, denoted by the term "fine structure," is believed to have an important effect on physical properties.

In this work, x-ray diffraction techniques have been applied for the quantitative measurement of parameters related to fine structure. Studies of relations among the various structural features and physical properties have also been undertaken.

EXPERIMENTAL TECHNIQUE

Equipment

The x-ray patterns were taken with a General Electric XRD unit using nickel-filtered copper

* Contribution No. 202 from the Chemical Department, Experimental Station.

¹R. O. Herzog and W. Jancke, *Zeits. f. Physik* 3, 196 (1920); J. Hengstenberg, *Zeits. f. Krist.* 69, 271 (1928); H. Mark and K. H. Meyer, *Zeits. f. physik. Chemie* B2, 115 (1929); H. Mark, *J. Phys. Chem.* 44, 764 (1940); W. A. Sisson, *Chem. Rev.* 26, 196 (1940).

radiation or in some cases monochromatized copper $K\alpha$ radiation. All patterns were taken with a 5-cm specimen-to-film distance. The films were processed under controlled conditions and the parameters were measured with a Photovolt photometer which had been remodeled and calibrated directly in x-ray intensity units.

Yarns Studied

With the exception of some saponified cellulose acetates, most of the samples studied were prepared by a conventional viscose process. Linters

TABLE I. Summary of x-ray parameters for viscose rayon.

Parameter	Description Orientation parameters—measured from azimuthal trace
$180/A$	A is the angular breadth in degrees of the interference at half maximum intensity, corrected for background intensity.
O , orientation number	Ratio of equatorial to polar intensity of the interference.
Orientation distribution	Angular distribution of the interference around the fiber axis, obtained by calculating the fraction of the total interference at each angle.
	Lateral order parameters—measured from radial trace
Reciprocal interference width (mm^{-1}) ($1/0.5W_s$)	Reciprocal of the half-width of the 101 interference at half-maximum intensity, corrected for background intensity, measured toward the 101 interference.
Radial intensity Ratio $\frac{(I_1 - I_m)}{I_1} \times 100$	I_1 —intensity of the 101 interference. I_m —minimum intensity between the 101 and 101 interferences.

cellulose was steeped in sodium hydroxide, aged, then reacted with carbon bisulfide to form the soluble xanthate. The xanthate was dispersed in alkali then extruded into a primary bath which contained sulfuric acid and sodium sulfate and in some cases zinc sulfate or zinc and iron sulfates. The coagulated and partially regenerated filaments from the primary bath were then passed into a secondary bath containing hot water and stretched. Large differences in properties were achieved by variations in the viscose or baths.

X-Ray Parameters

Orientation

Figure 1 illustrates the x-ray diffraction pattern and photometric traces for a viscose yarn which was spun into an acid-sulfate bath and highly stretched. Parameters frequently measured from the patterns are tabulated and described in Table I and discussed in more detail in the following. The photometric trace may be made along the arc of an interference, giving the symmetrical intensity *vs.* angle curve shown,

DIFFRACTION PATTERN

PHOTOMETRIC TRACE

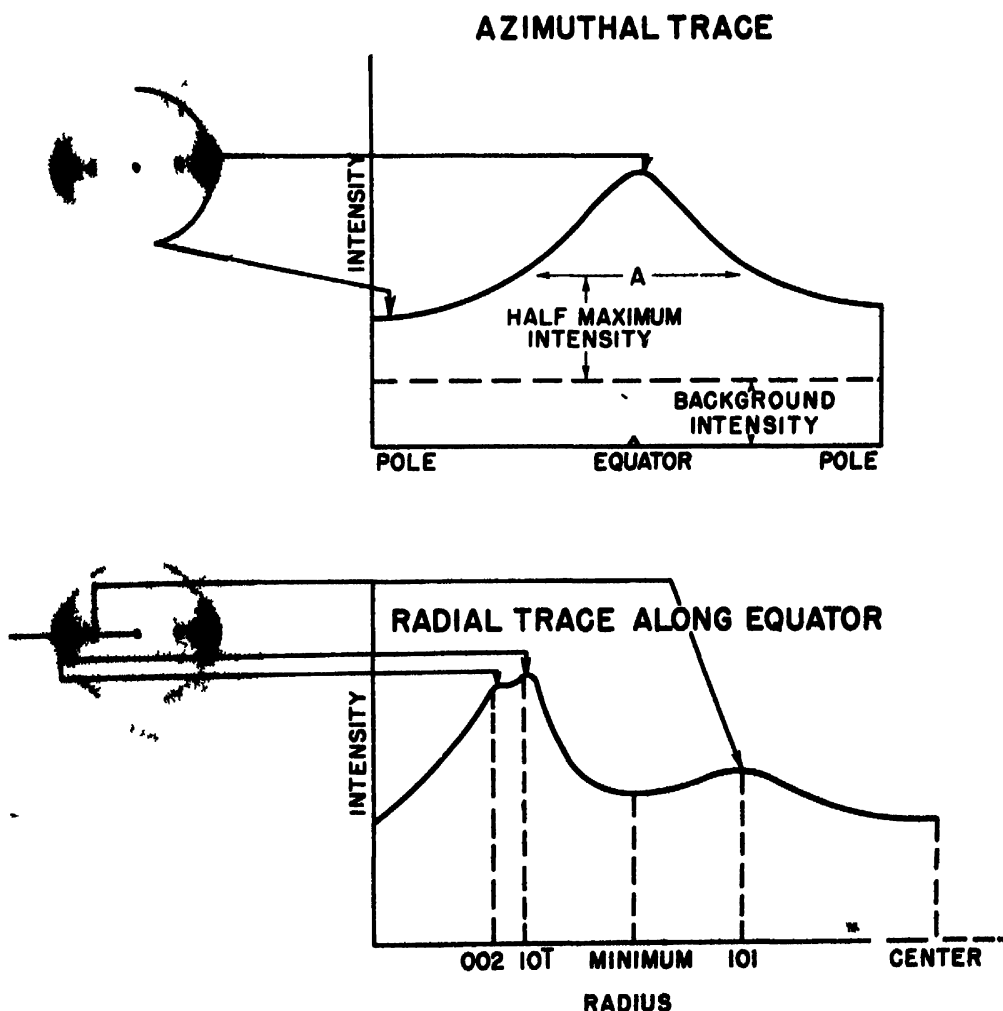


FIG. 1. Method of photometering x-ray diffraction patterns of viscose rayon.

which is a measure of the orientation of the cellulose crystallites. Complete orientation distributions can be calculated from this curve by procedures developed by Kratky, Hermans, Sisson, and others.³ The distribution represents the fraction of the interference intensity having any designated orientation and is particularly useful in studying small differences in orientation or high degrees of orientation, as well as differences in orientation distribution. The angular breadth of the photometric curve at half-maximum intensity, a parameter used by Kratky, Sisson, and others,³ is useful for measuring intermediate degrees of orientation and is more convenient to use than the complete distribution. We have expressed the measurements as 180° divided by the angle in order to provide a parameter which increases with orientation. We have also used another simple parameter, the ratio of the equatorial to the polar intensity, which is more suitable than the orientation angle for measuring low degrees of orientation.

Orientation of the long axes of cellulose crystallites parallel to the long axis of the fiber but with other crystallographic directions at random has been termed uniaxial orientation by Sisson.⁴ In this case, equivalent orientation parameters are obtained from the three principal interferences for cellulose since all three interferences arise from planes parallel to the long axes of the molecules. Other types of orientation arise when other planes or directions in the cellulose crystallites are aligned with respect to particular reference planes or directions of the sample. A special type of orientation frequently encountered in cellulose is orientation of the 101 diffraction planes parallel to a surface of the sample, referred to as selective uniplanar orientation by Sisson.⁵ In fibers this is manifested as orientation of the 101 planes parallel to the fiber surface (or axis). It is revealed by higher orientation parameters for the 101 interference than for the $10\bar{1}$ or 102 interferences. Unless otherwise

specified the term orientation as used here refers to uniaxial orientation.

Background Correction

A troublesome experimental detail which should be mentioned is the background scattering correction which must be used in photometry because of the considerable diffuse scattering observed in patterns of regenerated cellulose. In many yarns, for example, the orientation angles and radial widths cannot be measured at all before a background correction has been applied. This has been discussed in detail by Kratky and others,⁶ but no completely satisfactory solution has been arrived at. In the present work it was observed that cellulose yarns of low orientation frequently had a high degree of selective uniplanar orientation, that is, the 101 planes were more highly oriented than the $10\bar{1}$ or 002 planes. This selective orientation disappeared as the yarns were stretched and in highly oriented yarns the 101 planes actually appeared to be less highly oriented than the $10\bar{1}$ or 002 planes. There appears to be no physical reason for the 101 planes to develop a lower orientation than the other lateral planes at high uniaxial orientation. Consequently, it was assumed that after suitable background correction the orientation parameters measured from all three interferences should be equal in the highly oriented yarns. The minimum intensity between the 101 and $10\bar{1}$ interferences at 45° from the equator was chosen as an arbitrary correction factor since it was found to give about equal angles for all three interferences over quite a range of orientations. Although the structural significance of such a correction is uncertain, it serves the useful practical purpose of broadening the range of samples for which interference angles and half-widths can be measured.

Lateral Order

The photometric trace may also be taken radially along the equator of the pattern, Fig. 1, to obtain an intensity-distance curve representing the interferences arising from the three principal lateral spacings in crystalline cellulose. From this curve the intensities, positions, and

³ O. Kratky, *Zeits. f. Electrochemie* **48**, 587 (1942); P. H. Hermans, *Rec. trav. chim. des Pays-Bas* **63**, 13-24 (1944); W. A. Sisson and G. L. Clark, *Ind. Eng. Chem. Anal. Ed.* **5**, 296 (1933).

⁴ P. H. Hermans, O. Kratky, and R. Treer, *Kolloid Zeits.* **96**, 30 (1941); W. A. Sisson, *Textile Research J.* **425** (1937).

⁵ W. A. Sisson, *J. Phys. Chem.* **40**, 343 (1936).

⁶ W. A. Sisson, *J. App. Phys.* **44**, 523 (1940).

⁶ P. H. Hermans, O. Kratky, and P. Platzek, *Kolloid Zeits.* **86**, 248-52 (1939).

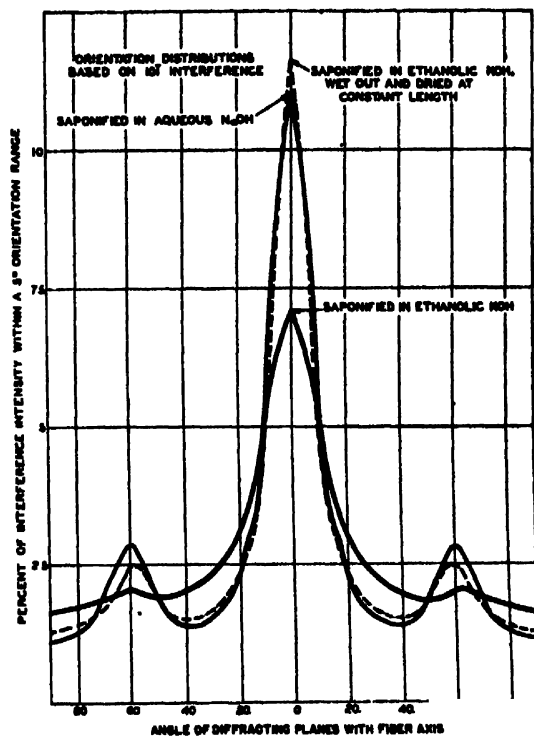
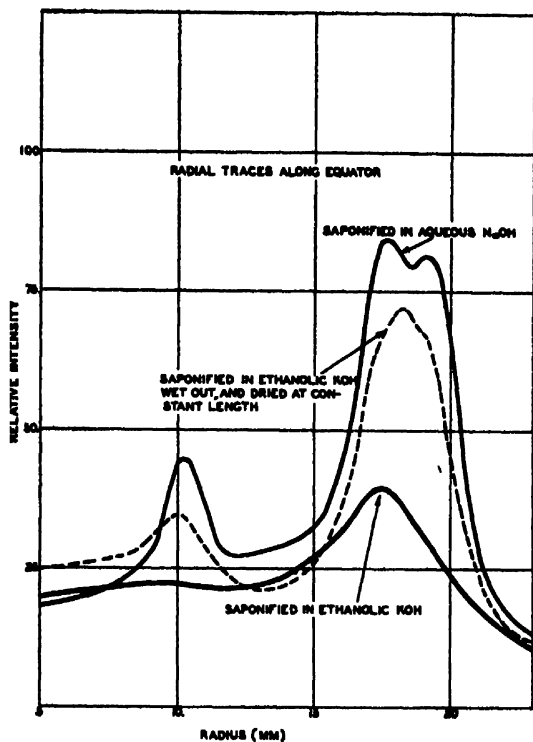


FIG. 2. Lateral order and orientation of cellulose yarns regenerated in aqueous and non-aqueous media.

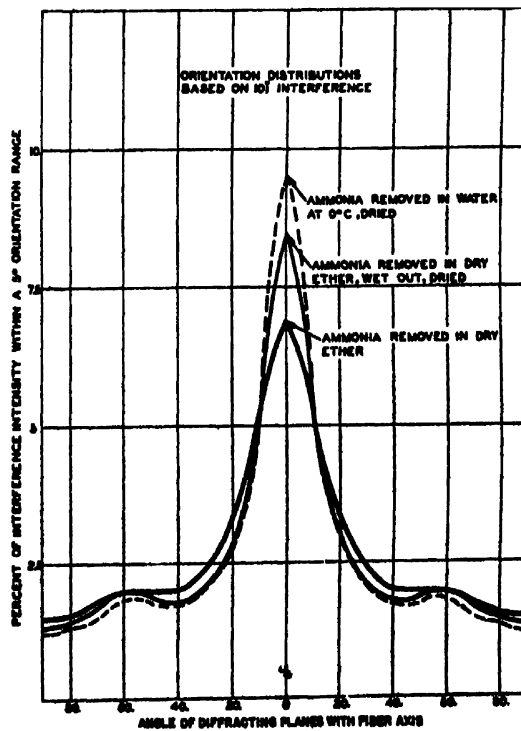
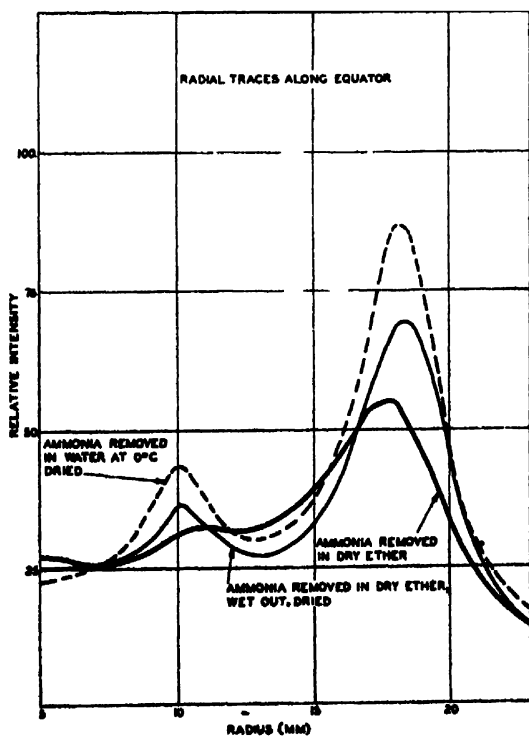


FIG. 3. Lateral order and orientation of cellulose yarns after swelling-deswelling treatments.

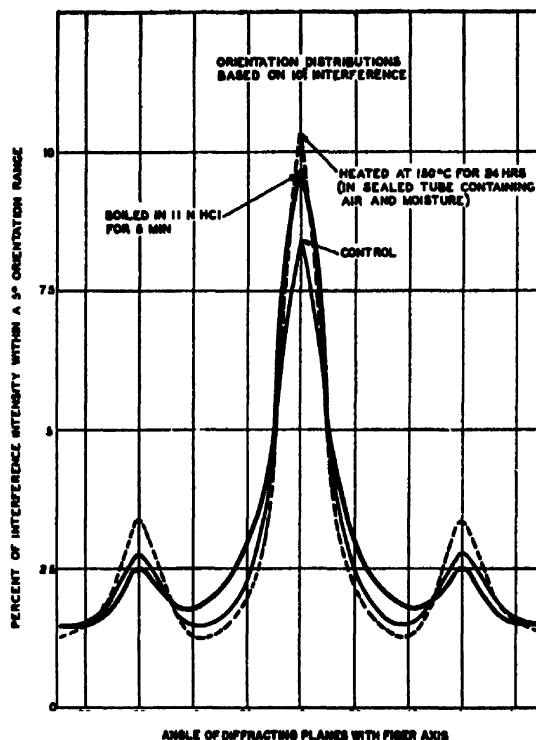
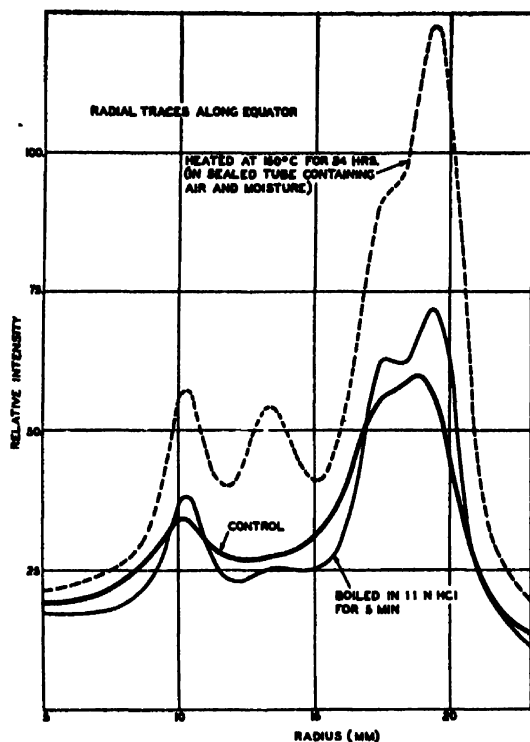


FIG. 4. Lateral order and orientation of chemically degraded cellulosic yarns.

radial widths of the interferences can be measured. The widths, shapes, and intensities of the interferences are influenced by the dimensions and perfection of ordered regions and also by other fine structural features such as the amount, distribution, and arrangement of ordered and disordered regions. These structural features for the lateral planes depend largely upon the local disordering of the polymer chains in the direction perpendicular to their long axes, referred to as lateral disordering by Baker and Fuller.⁷ This terminology is followed here, the width and intensity measurements being referred to merely as parameters related to lateral order in view of the uncertainty of more detailed structural interpretation for the very diffuse and overlapping interferences from some regenerated celluloses. The positions of the interferences are not used in this paper since the average interplanar spacings for the more ordered cellulose structures studied differ very little from those for quite disordered structures.

The term lateral order is thus intended here to emphasize the fact that the x-rays give an overall picture of the degree of regularity of the lateral diffracting planes. Most of the regenerated celluloses here studied appear to be particularly unfavorable for attempts to assign specific structural causes such as variations in "degree of crystallinity" for the observed planar irregularity. Lateral ordering is of course to be differentiated from longitudinal ordering, i.e., ordering along the long chain axes.⁷ Longitudinal ordering, which largely determines the sharpness of the non-equatorial interferences, is not dealt with in this paper.

EXPERIMENTAL RESULTS

Relation of Lateral Order to Orientation

Regenerating or Deswelling in Non-Aqueous Media

It has been noted that "crystallinity" as well as the orientation of cellulosic materials increases during stretching.⁸ In considering the application

⁷ W. O. Baker and C. S. Fuller, J. Am. Chem. Soc. 65, 1120 (1943).

⁸ P. H. Hermans, Proc. Acad. Wetenschappen Amsterdam 43, 1032 (1940); Kolloid Zeits. 96, 311 (1941).

of quantitative measurements of these features, the question then arises as to whether orientation may vary independently of lateral order. Sisson has observed that celluloses of very low order are produced by regenerating a cellulose derivative in non-aqueous media or by removing an intracrystalline swelling agent with a solvent less polar than water.⁹ These effects have been examined quantitatively in order to indicate some of the possible relations between changes in lateral order and changes in orientation.

Figures 2 and 3 show photometric traces of diffraction patterns for cellulose yarns prepared from cellulose acetate by saponification in aqueous and non-aqueous media and for cellulose yarns after swelling in liquid ammonia and deswelling in aqueous and non-aqueous agents. The equatorial traces show that the interferences for the samples deacetylated or deswollen in aqueous media are very sharp and intense, indicating well-ordered diffracting planes. On the other hand, the interferences for the samples treated in non-aqueous media are weak and diffuse and the 101 interferences appear only as small inflections in the traces, showing that the cellulose chains are not sufficiently ordered to form well-defined diffracting planes. The chains are now nearly randomly turned about their long axes so that lateral separation of the chains is approximately the same in all directions, in contrast to the three quite different principal lattice spacings shown by a well-ordered cellulose. The orientations of the deswollen yarns, Fig. 3, are very little different from those of the original, which is the yarn of intermediate orientation in Fig. 7.

The samples prepared in non-aqueous media were unstable in that merely wetting and re-drying markedly increased lateral order. Removal of the substituent groups or swelling agent in non-aqueous media leaves an irregular structure evidently because such media do not soften the cellulose sufficiently to allow ordering during the transition from the substituted or swollen network to the higher density cellulose. Wetting with water provides the necessary mobility and the chains rearrange, greatly increasing lateral

order. The increase in order on wetting was accompanied by a marked increase in orientation in the case of the saponified cellulose acetate but there was not much change in orientation on wetting deswollen yarn.

It will be noted that wetting out the yarn deswollen in non-aqueous media converted it to a structure nearly as well ordered as that of the original untreated sample, shown as the yarn of intermediate orientation in Fig. 7. Comparable results were obtained by applying this swelling-deswelling and wetting out treatment to a yarn which was originally much more highly ordered as shown by the sharper radial trace in Fig. 7. The maximum degree of ordering achievable by drastic swelling followed by deswelling and aqueous annealing thus appears to be limited by the structure of the yarn originally set up during spinning. This point will be discussed further in a later section.

Mild Chemical Degradation

Another example of change in lateral order is furnished by chemical degradation. Mild chemical degradation of cellulose can greatly increase lateral order with substantially no change in orientation and no perceptible change in crystal

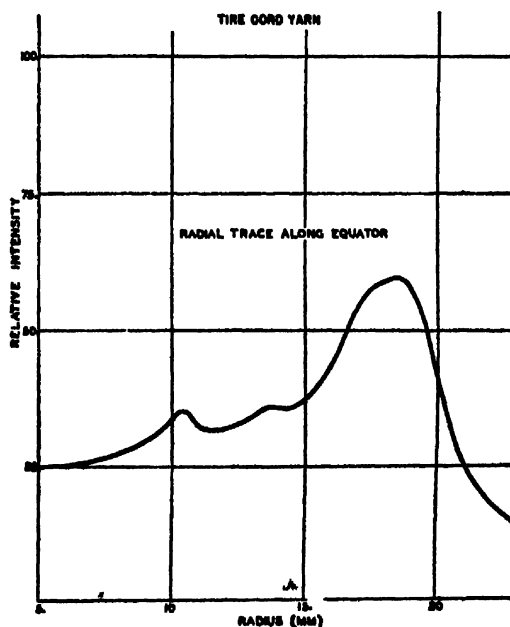


FIG. 5. X-ray diffraction from a yarn containing both mercerized and native cellulose crystal structures.

⁹ W. A. Sisson, *Ind. Eng. Chem.* **30**, 530 (1938); W. A. Sisson, *Cellulose and Cellulose Derivatives* (Interscience Publishers, New York, 1943), first edition, Chapter IIIA.

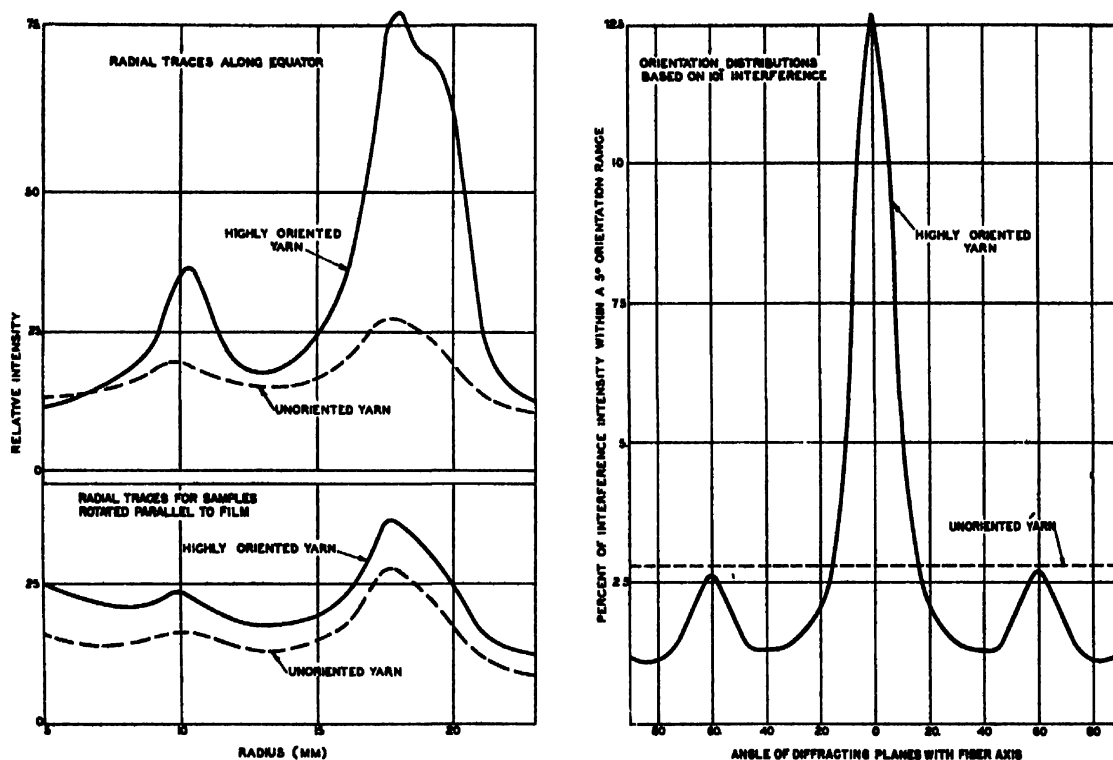


FIG. 6. Effect of orientation on the measurement of lateral order

structure. Figure 4 shows the photometric traces of a highly stretched tire cord yarn before and after 5 minutes boiling in 1.1 normal hydrochloric acid and after 24 hours heating in the presence of air and moisture. It is known that either of these treatments can cleave glucosidic linkages in the cellulose chains. This process occurring preferentially in the disordered, more accessible regions of the sample might release these less ordered chain segments sufficiently to allow rearrangement, and thus cause the over-all increase in lateral order which is observed.

Comparison of Cellulose with Other Polymers

The preceding examples show that lateral order can, to some extent, vary independently of orientation, a phenomenon which has been reported for other polymeric systems such as polyamides and cellulose esters by Baker.¹⁰ The importance of distinguishing between changes in lateral order and orientation is emphasized. The increase in ordering on wetting the metastable cellulose obtained by regenerating or de-

¹⁰ W. O. Baker, *Ind. Eng. Chem.* **37**, 246-54 (1945).

swelling in non-aqueous media appears to be entirely analogous to the effects of annealing polyamides reported by Fuller, Baker, and Pape¹¹ and effects of heat and solvents on cellulose esters, described by the same authors.¹²

Lateral Order Parameters for Regenerated Cellulose

The differences in lateral order just illustrated can be expressed quantitatively in terms of the interference widths or intensities which can be measured from the photometric traces shown in Fig. 1. Unfortunately, the 101 interference is too diffuse in many rayons for accurate width measurement. The intensity of the 002 interference falls off rapidly, proceeding from the maximum outward along the radial trace, and furthermore, the position of the maximum is hard to determine because of the doublet formation with the 10 $\bar{1}$ interference. Consequently the measured widths of the 002 interference vary little for large

¹¹ W. O. Baker, C. S. Fuller, and N. R. Pape, *J. Am. Chem. Soc.* **62**, 3275-81 (1940).

¹² W. O. Baker, C. S. Fuller, and N. R. Pape, *J. Am. Chem. Soc.* **64**, 776-82 (1942).

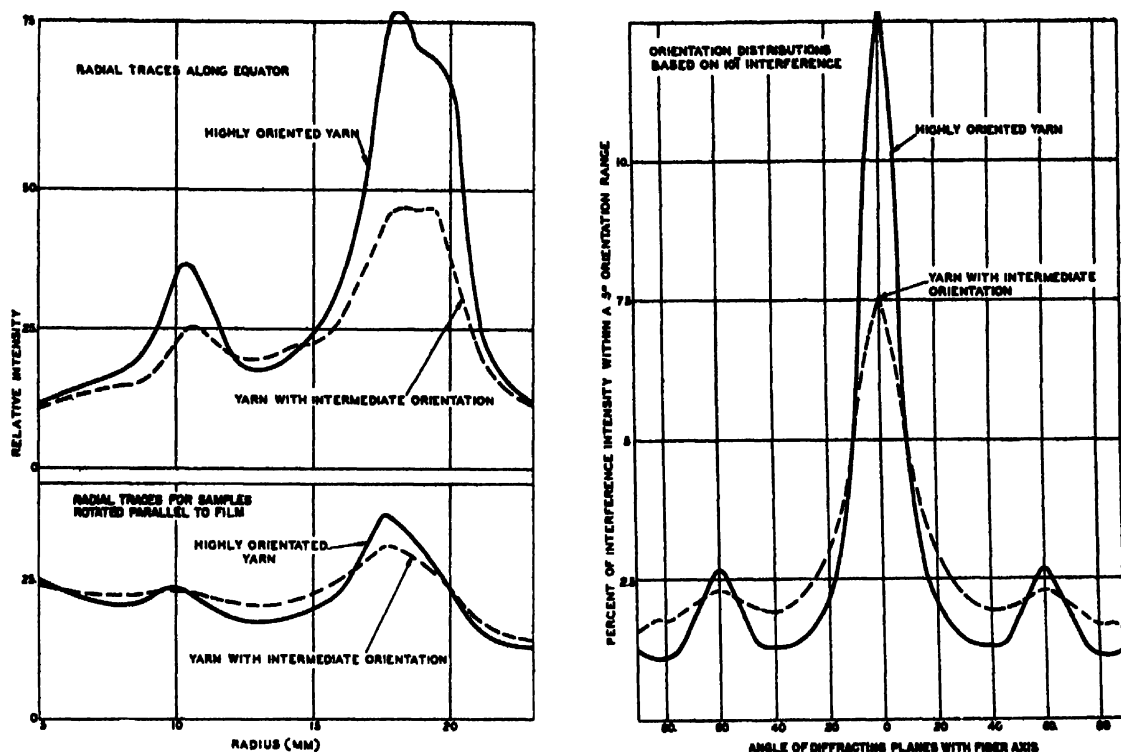


FIG 7 Effect of orientation on the measurement of lateral order.

changes in pattern. The $10\bar{1}$ interference width can be measured only toward the center of the pattern because of the $10\bar{1}$ -002 doublet formation and its usefulness is limited also by the fact that the native cellulose 101 - $10\bar{1}$ doublet occurs just inside the $10\bar{1}$ interference for regenerated cellulose, Fig. 5. Most rayon samples contain detectable amounts of native cellulose structure and many contain large proportions of it. A large difference in native cellulose content could change the $10\bar{1}$ interference width even though there is no difference in lateral order. Nevertheless this parameter serves in many instances as a practical measure of pattern diffuseness. The width is expressed as a reciprocal in order to provide a parameter which increases with increasing lateral order.

X-ray intensities directly measured from the photographic patterns were not sufficiently reproducible at times because of variations in x-ray dosage. Consequently, an intensity ratio was adopted as an expedient for measuring differences in pattern sharpness more reproducibly. It was noted that the 101 interference almost

disappears for celluloses of low lateral order and shows a large range of intensity depending upon sample history. The height of the 101 interference above the minimum between the 101 and $10\bar{1}$ interferences, expressed as a percentage of the 101 intensity, $[(I_1 - I_m)/I_1]100$, provides a parameter which increases with lateral order. This empirical ratio, which will be referred to as the radial intensity ratio, can be measured fairly conveniently and reproducibly by photometric methods.

Effect of Orientation on Lateral Order Parameters

It is emphasized that these lateral order parameters are only comparable if the samples have the same orientations. Small differences in orientation have little effect but large differences change the background corrections, interference intensities, and the contributions from non-equatorial interferences, with accompanying effects on the intensity ratios and reciprocal width parameters just described. In addition to this effect of uniaxial orientation, selective uniplanar orientation, that is, higher orientation of

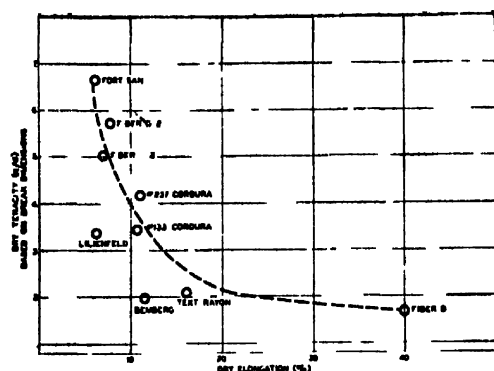


FIG. 8. Tenacity vs. elongation for standard rayons

the 101 planes with respect to the $10\bar{1}$ or 002 planes, will increase the 101 intensity relative to the minimum. This produces an increase in intensity ratio even though there is no change in lateral order. Also, in samples containing appreciable amounts of material with the native cellulose crystal structure, special corrections must be made for the effect of native cellulose interferences on the intensity ratio. These latter two difficulties will be further discussed later as encountered in specific cases. To illustrate the effect of uniaxial orientation, Fig. 6 shows photometric traces of two yarns which have much different orientations and apparently different lateral order as revealed by the radial traces along the equator. If the yarn specimen is rotated in a plane perpendicular to the x-ray beam, during exposure, the interferences are distributed on uniform rings and the radial trace is equivalent to a radial trace averaged from equator to pole in the original unrotated patterns. The average traces, obtained from rotated samples, show that these two yarns actually have quite similar lateral order, the difference originally noted arising largely from the difference in orientation. Figure 7 reproduces the traces for samples differing widely in orientation and also apparently in lateral order, when the estimate is made from the original differently oriented samples. X-ray diagrams of these same yarns, averaged by rotating the samples, demonstrate that they are still quite different with respect to lateral order, even when compared at the same orientation.

It should be noted that rotating the samples does not completely average the diffraction,

since the orientations of the various planes are not randomized by rotating around one axis. This discrepancy in comparing the lateral order is of course more important where the differences in orientation are largest. An attempt was made to randomize the samples completely before x-raying by cutting into lengths roughly equal to their diameter. However, a serious difficulty was encountered here in that it was difficult to be sure that the cutting had not altered fine structure. It was found, for example, that room temperature cutting considerably decreased lateral order as did also rolling dry cellulose filaments. This effect appears to be closely comparable to the disordering of cellulose derivatives by rolling described by Baker¹⁰ and to the disordering of cellulose during grinding reported by Hess and co-workers,¹² indicating that mechanically working an unswollen cellulose decreases its lateral order. Cutting the cellulose at very low temperatures, such as dry ice temperature, appears to reduce the concurrent changes

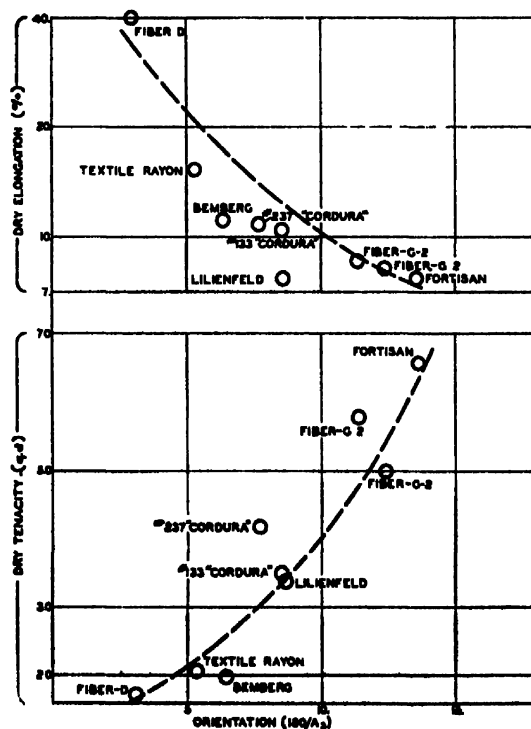


FIG. 9. Tenacity-elongation-orientation relations for standard rayons.

¹² K. Hess, E. Steurer, and H. Fromm, *Kolloid Zeits.* [1] 98, 148-159 (1942); [2] 98, 290-304 (1942).

in fine structure, but it is not certain that even this method of randomization is satisfactory.

These averaging procedures tend to increase the relative background scattering intensity and thus add to the difficulties of measuring small differences. However, the results show that a comparison of the lateral order of samples differing widely in orientation can be very misleading unless some correction is made for the difference in orientation.

Correlation of Properties with Fine Structure

The structural features, orientation, and lateral order, are considered to have important effects on physical properties of cellulose. It is well known from the work of Herzog, Lilienfeld, Berkley, Morey, Sisson, and others that many physical properties, including tenacity and elongation, vary in a regular way with orientation.¹⁴ Hermans states that the properties and quality

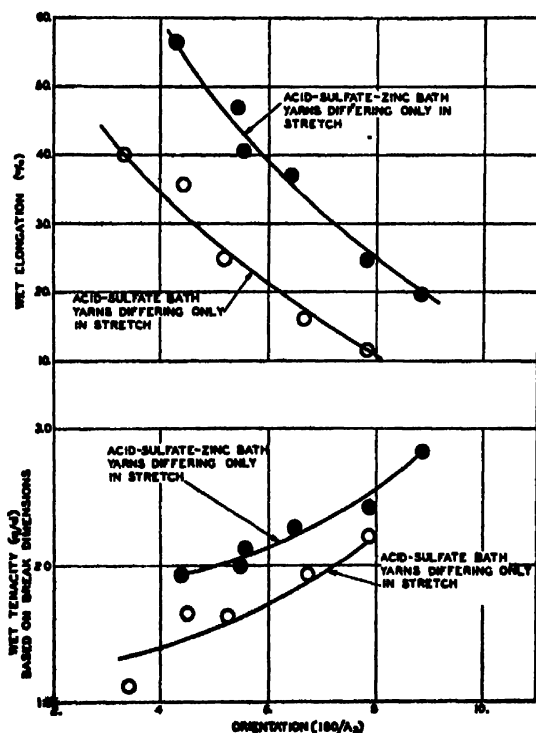


FIG. 10. Wet tenacity-wet elongation-orientation relations for experimental rayons.

¹⁴ R. O. Herzog, *J. Phys. Chem.* **30**, 457 (1926); L. Lilienfeld, *Brit. Patents* 274,521 (Jan. 11, 1926); 274,690 (Jan. 11, 1926); E. E. Berkley, *Textile Research* **9**, 355 (1939); D. R. Morey, *Textile Research* **4**, 491 (1934); W. A. Sisson, *Textile Research* **7**, 425 (1937).

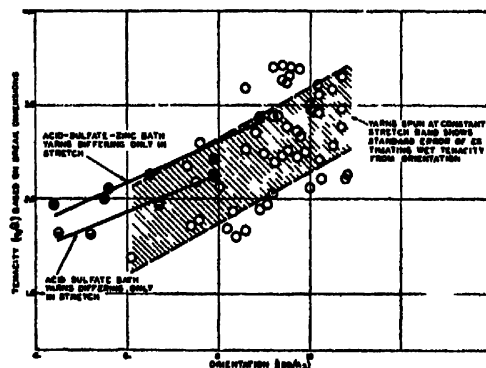


FIG. 11. Wet tenacity vs. orientation for experimental rayons.

of regenerated fibers depend upon (1) the degree of polymerization and distribution of chain length; (2) the preferred orientation of the crystallites and molecules and the distribution of this orientation; (3) the ratio of crystalline to non-crystalline material; and (4) the average distribution of the above over the cross section of the fiber.¹⁵

Mark has proposed that flexibility and reactivity are dependent upon the disordered regions in cellulose while tenacity and elastic modulus are related to the amount of ordered material present.¹⁶ In the present work quantitative x-ray measurements have been applied to ascertain whether this effect of disorder on properties can be verified experimentally.

Tenacity-Elongation-Orientation Relations for Standard Rayons

As a broad introduction to the data, Fig. 8 shows tenacities plotted against elongations for a number of commercial types of cellulose yarns. Tenacity decreases with increasing elongation in such a way that a smooth curve averaging the points has a hyperbolic shape. This relation is well known, being indicated by the rough constancy of the product of tenacity and elongation which has been called the "silk factor" or "work factor."¹⁷

This tenacity-elongation relationship for yarns differing widely in orientation results from the fact that both tenacity and elongation vary with

¹⁵ P. H. Hermans, *J. Phys. Chem.* **45**, 827 (1941).

¹⁶ H. Mark, *J. Phys. Chem.* **44**, 764 (1940).

¹⁷ D. Smith, *Proc. A.S.T.M.* **44**, 543 (1944).

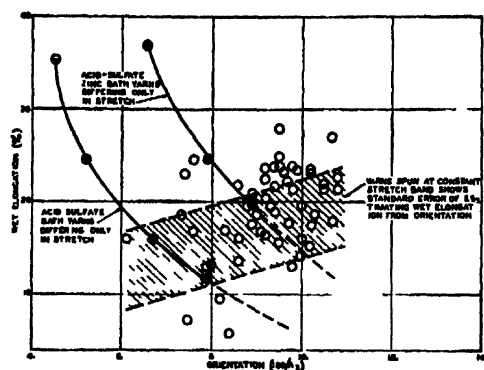


FIG. 12. Wet elongation vs orientation for experimental rayons¹⁸

the orientation of cellulose chains. This dependency is illustrated in Fig. 9 where both tenacity and log elongation are plotted against orientation for these yarns. Tenacity appears to increase substantially linearly with increasing orientation while elongation decreases hyperbolically with increasing orientation. The increase in tenacity with orientation has been explained qualitatively by Mark on the basis that the more highly oriented materials have more crystallites aligned parallel to the fiber axis and hence in a position to support the load.¹⁸ Elongating the filaments increases the orientation so that if all samples break at complete or comparable orientations, more elongation would be possible in a less highly oriented sample. Preston has calculated elongation on this basis from refractive index measurements of orientation and obtained qualitative agreement with the observed values.¹⁹

The scatter of the data in Fig. 9 indicates that factors other than orientation also affect these properties. The curves shown have been drawn to represent the highest tenacities and elongations so far observed at each orientation. From the measured orientation of a given yarn and these reference curves, one can estimate the tenacity and elongation which the yarn should develop if it is as efficient as the reference yarns in converting orientation into these physical properties. The ratios of estimated to observed properties, which may be termed "orientation

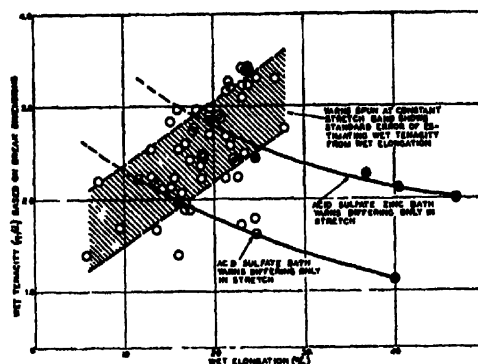


FIG. 13. Wet tenacity vs wet elongation for experimental rayons.

efficiencies," serve as a convenient indication of the effects of factors other than orientation and have been helpful in evaluating new yarns and spinning processes. In some instances, for example, low efficiencies have been traced to mechanical non-uniformities of the yarns. In other cases it has been possible to show that some structural features other than orientation were important.

Tenacity—Elongation—Orientation Relations for Experimental Rayons

Wet tenacity-orientation, wet elongation-orientation curves for two experimental yarns at various degrees of stretch are shown in Fig. 10. One yarn was spun into an acid-sulfate bath, the other into an acid-sulfate-zinc bath. Dry tenacities and elongations plotted against orientation for these same yarns at various stretches give curves which are quite similar but in general not as smooth as those shown for the wet properties. Differences between dry and wet properties, the treatment of which is beyond the scope of this paper, may be neglected for the purposes of this discussion. The linear tenacity-orientation and hyperbolic elongation-orientation relation indicated roughly by data for several types of rayon in Fig. 9 are shown more clearly here for these yarns in which the orientation is increased by merely increasing spinning stretch. The curves for these two experimental yarns have much the same form and slopes but are displaced from each other because of the effects of variables other than orientation. Changes in bath or viscose variables other than adding zinc

¹⁸ H. Mark, *Melliand Textiber* 10, 695 (1929); *Pap. Trade J.* 113, 34 (July 17, 1941).

¹⁹ J. M. Preston, *Modern Textile Microscopy* (Emmott and Company, London, 1933), first edition, p. 265.

to the bath have produced a similar effect, forming families of tenacity-orientation and elongation-orientation curves, respectively. At any given orientation, then, a range of tenacities and elongations is possible, depending upon other factors, as shown by Sisson.⁹

Some information on the effects of fine structural factors other than orientation was obtained by examining a series of yarns spun by identical mechanical procedures. Large changes in properties of these yarns were achieved by changing the acidity or metal ion concentration in the coagulating bath or the degree of substitution, salt index, or composition of the viscose. All samples were collected on a bobbin and dried at constant length. These changes in spinning conditions and viscose produced large differences in lateral order and marked differences in orientation, even though secondary bath stretch was held constant. The graph of wet tenacity against orientation for these yarns, Fig. 11, shows that the data scatter considerably, suggesting that variables other than orientation are important. There is, nevertheless, a trend for wet tenacity to increase with orientation and the over-all linear correlation coefficient calculated for this relation has high statistical significance. The shaded band shown represents the standard error of estimating wet tenacity from the best straight line expressing tenacity as a function of orientation, calculated by the method of linear regression.²⁰ The increase in wet tenacity with orientation brought about by stretch alone shown in Fig. 10, is also reproduced in the solid curves of Fig. 11 for comparison. Differences in orientation apparently have similar effects on tenacity whether achieved by variations in viscose and bath or by differences in degree of stretching.

With elongation, however, the situation appears superficially to be quite different from that encountered when stretch is varied. Elongation now appears to increase somewhat with orientation, Fig. 12. The solid curves show for comparison the wet elongation-orientation relations previously observed for yarns stretched different amounts, Fig. 10. The scatter of this wet elongation-orientation plot is also very large, again

indicating that variables other than orientation are important.

Since both wet tenacity and wet elongation increase with orientation for these yarns spun at constant stretch, it follows that wet tenacity must also increase with wet elongation, as illustrated in Fig. 13. The solid curve represents the previously shown hyperbolic wet tenacity-wet elongation curve for yarns differing mainly in orientation. The over-all increase in wet tenacity with wet elongation encountered here suggests that the stress-strain curves may have generally similar shapes, the yarns merely breaking at different elongations. Obviously some specific yarns as shown on the slide must have considerably different stress-strain curves because of the observed large spread in elongation at a given tenacity or in tenacity at a given elongation.

Statistical Analysis of the Correlations among Structure and Property Parameters

In general, there appear to be relations among wet tenacity, wet elongation, lateral order, and orientation of these yarns, but the data scatter considerably, indicating that variations in other factors are important. Mathematical procedures have therefore been applied for a quantitative study of the complex interrelationships among the variables. Applying the usual statistical methods, linear correlation coefficients were calculated for each of the six possible pairs of variables, Table II. Statistical significance of

TABLE II. Total and partial linear correlation coefficients for properties and x-ray parameters of experimental rayons.

$n = 55$ Variables Wet tenacity (based on break dimensions) Wet elongation Orientation $(180/A_s)$ Lateral order $[(I_1 - I_m)/I_1] 100$					
Variables correlated	Total coefficients	"t" factors	Variables held constant for partial coefficients	Partial coefficients	"t" factors
Tow-O	+0.5312*	4.48	EW, L	+0.4034*	3.15
EW-O	+0.3910*	3.04	Tow, L	+0.0804	0.58
Tow-EW	+0.5891*	5.21	O, L	+0.4310*	3.41
EW-L	-0.5496*	4.71	Tow, O	-0.4492*	3.59
Tow-L	-0.3478*	2.65	EW, O	-0.0133	0.09
L-O	-0.2595	1.92	Tow, EW	-0.0477	0.34
"t" factor required for significance					
5% level					
1% level					

* M. J. B. Ezekial, *Methods of Correlation Analysis* (John Wiley and Sons, New York, 1941), second edition.

* Coefficient considered statistically significant.

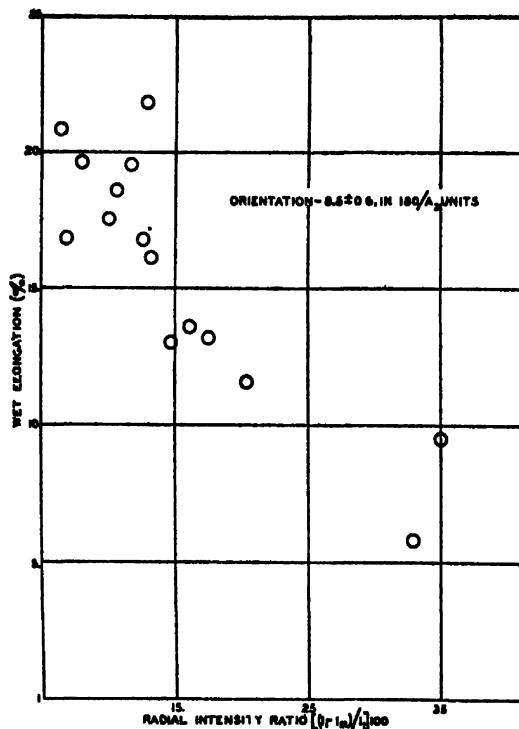


FIG. 14. Wet elongation vs. lateral order for experimental rayons of constant orientation.

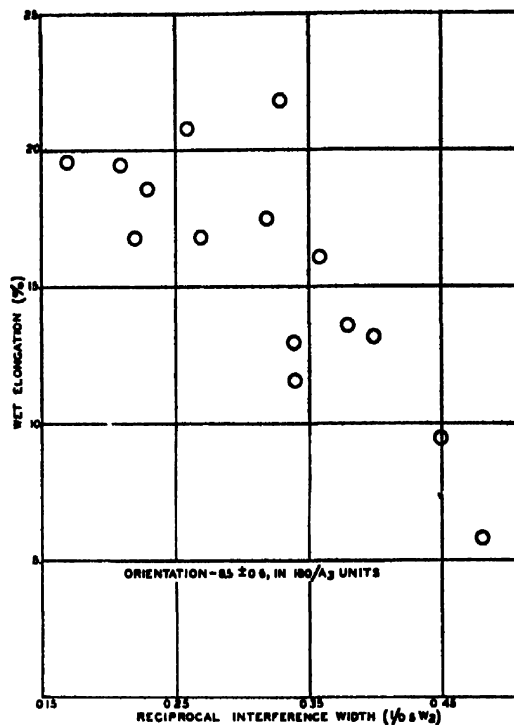


FIG. 15. Wet elongation vs lateral order for experimental rayons of constant orientation.

these coefficients has been estimated by use of the familiar "Student" *t*-function. Significant positive correlation coefficients are found for the pairs of variables wet tenacity-orientation, wet elongation-orientation, and wet tenacity-wet elongation. These are expected from the correlation plots which were illustrated in Figs. 11-13. An especially interesting feature of the entire group of coefficients, however, is the fact that they indicate that lateral order as well as orientation has a very important effect on the physical properties of these yarns. Both wet tenacity and wet elongation are found to decrease significantly with increasing lateral order.

In each correlation, however, the specific relations sought are obscured by the fact that other variables have not been held constant. For example, both wet tenacity and orientation are known to be correlated with wet elongation and the observed coefficient for wet elongation with lateral order has undoubtedly been affected by the accompanying changes in wet tenacity and orientation.

In order to isolate the specific effect of a given

structural parameter on a given property it is necessary to hold the other variable constant. This can be done mathematically by the methods of partial correlation which have been worked out by Ezekiel and others.²⁰ Linear relationships have been assumed between all six pairs of variables here since this greatly simplifies the calculation and since more complicated procedures have not so far appeared justified because of the scatter of the data.

When orientation and wet tenacity are maintained constant by the method of partial linear correlation, the relation between wet elongation and lateral order remains statistically significant, Table II. However, at constant wet elongation and orientation, wet tenacity is no longer significantly associated with lateral order, indicating that the apparent relationship shown by the total coefficient was merely the result of simultaneous variations in wet elongation and the dependence of wet tenacity upon wet elongation. The partial correlation coefficients also show that wet tenacity is correlated with orientation even at constant wet elongation and lateral

order, as might be expected. Wet tenacity is still significantly related to wet elongation even when orientation and lateral order are maintained constant, indicating that variables other than orientation and lateral order are important here. This correlation again suggests the similarity of stress-strain curve previously discussed and raises the question of what factors cause these yarns to fail at different elongations. Useful information concerning this problem might be obtained by studying the entire stress-strain curves rather than merely the properties at break. It is possible that the structure of some samples will correlate more closely with the elastic modulus than with tenacity and elongation at break.

When wet tenacity and lateral order for these samples are maintained constant, wet elongation is not significantly related to orientation. This indicates that the relationship observed between these two variables when wet tenacity and lateral order vary is merely the result of the close association between wet tenacity and wet elongation.

The Elongation-Lateral Order Relationship

The wet elongation/lateral order relation can be demonstrated graphically as well as by the analytical procedure just described. Orientation was maintained constant in Fig. 14 by merely plotting the data for all yarns having essentially the same orientation. The group was taken from the lower orientation end of the samples, where the differences in elongation tended to be larger. The close correlation between lateral order and wet elongation at constant orientation is clearly shown. Although lateral order as revealed by the radial intensity ratio has been used here, essentially the same correlation is observed if reciprocal interference widths are used, Fig. 15. A similar relationship between dry elongation and lateral order was also found, although the data tend to scatter more. Reciprocal interference widths have sometimes been used to calculate crystallite dimensions for native cellulose.²¹ As mentioned previously, however, there are considerable uncertainties which make this spe-

cific interpretation of the x-ray data for a material such as regenerated cellulose appear of doubtful value.

Correction for Native Cellulose Content

Some of the changes in spinning conditions in producing the constant orientation samples just discussed also caused large changes in native cellulose structure content. The appearance of the native cellulose interferences in a typical pattern has already been illustrated in Fig. 5. As mentioned previously, the native cellulose can interfere seriously with determination of the lateral order parameters. In order to get some measure of the native cellulose content, the intensity ratio,

$$\left[\frac{I_{1n} - I_m}{I_1 + I_{1n} - 2I_m} \right] \times 100,$$

was calculated. I_{1n} is the intensity of the 101 native cellulose interference, I_1 is the 101 interference intensity for mercerized cellulose and I_m is the minimum intensity between the 101 and $10\bar{1}$ mercerized cellulose interferences, Fig. 5. Such a calculation is obviously crude because of the proximity of these interferences and their diffuse nature. However, lacking a better method, this procedure serves to grade samples according to apparent native cellulose interference intensity. The elongation-lateral order correlation can then be calculated just for the samples whose native cellulose intensity ratios fall within certain limits. An alternative correction method consists in adding a term for the native cellulose intensity ($I_{1n} - I_m$) to the numerator of the radial intensity ratio. With either correction procedure, however, the correlation between wet elongation and lateral order is still statistically significant.

Applicability of the Relationships

These quantitative measurements adequately confirm the long suspected influence of disordered regions in cellulose, proposed by Mark¹⁸ and Hermans,¹⁹ as previously described. Baker, Fuller, and Pape have shown that toughness and extensibility of cellulose fibers are favored by local chain disorder.¹² Baker points out that these phenomena are presumably general for all linear polymers,¹⁰ a conclusion which is strongly

²¹ H. Mark and K. H. Meyer, *Zeits. f. physik. Chemie* B2, 115 (1929).

supported by the present work on cellulose. It is emphasized, however, that these results do not mean that elongation of cellulose yarns will always increase as lateral order decreases. On the contrary, it would be expected that at constant orientation and lateral order other factors, which were nearly constant here, could operate to change the properties. There is some experimental evidence to be mentioned later that the properties of a yarn having the same orientation and lateral order as revealed by x-rays may vary according to the route followed in preparation. While some of these differences in properties may be caused by extraneous factors such as mechanical differences, it seems likely that structure also has had an important effect. For example, structural details not differentiated by x-ray diffraction, such as imperfections, dimensions, or amount of ordered regions, may have altered properties. The range of applicability and the limitations of this correlation between elongation and lateral order should become more clearly defined when information on structure and properties is available for yarns prepared under a wider variety of conditions.

Structural Changes During Spinning

Study of the yarn during each phase of preparation should be helpful in understanding the final structure and in accounting for properties.

Effect of Method of Coagulation and Regeneration

The timing of the coagulation and regeneration processes during spinning has an important effect on the lateral order and orientation of the final sample, as shown in Fig. 16. A yarn coagulated, stretched as the xanthate, then regenerated had much higher lateral order at the same orientation and could be oriented much more than a yarn which was simultaneously coagulated and partly regenerated before stretching. Regeneration occurring along with coagulation appears to fix a portion of the solution type disorder in the coagulating, shrinking structure. The regeneration occurring during coagulation allows the formation of interchain bonds which may restrict further ordering. Coagulation followed by regeneration evidently provides much more opportunity for ordering before the restricting cellulose

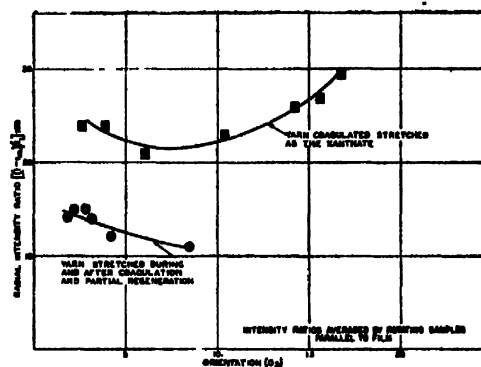


FIG. 16 Lateral order vs. orientation for yarns coagulated and regenerated by various methods.

interchain forces develop. It will be noticed that lateral order parameters for both yarns tend to decrease with increasing orientation in the low orientation range. The yarn stretched as xanthate could be oriented much more and at the higher orientation the order parameters pass through a minimum and finally increase with orientation. It is emphasized that orientation has an important effect on these relative lateral order parameters and it is not possible to decide from these data just how lateral order has changed with orientation; the data are closely comparable only at constant orientation. The lateral order parameters were calculated from patterns of rotated samples but this does not completely correct for differences in uniaxial orientation. The discrepancy is most important at the highest orientation and would make the order parameters too large. Both yarns show considerable selective orientation of the 101 planes at the lower orientations and this would also tend to increase the intensity ratio used as a measure of lateral order.

Skin-Cote Structure

Study of some of these yarns is further complicated by a structural discontinuity across the filaments, the so-called skin effect described by Preston, Sisson, and others.²² These investigators

²² J. M. Preston, *J. Soc. Chem. Ind.* 199-203 (June 12, 1931); K. Ohara and R. Mizuno, *Artificial Silk Stable Fiber J. Japan* 1, 10-12 (1933); K. Ohara, *Sci. Pap. Inst. Phys. Chem. Research (Tokyo)* 23, 152-8 (1934); W. Schramek and J. Helm, *Monatschr. Textil-Ind.* 52, 317-28 (1937); W. Schramek, *Papier-Fabrikant* 36, 547 (1938); W. A. Sisson, talk at Gibson Island Textile Conference, 1943.

have reported that the proportion of skin can be varied widely by suitable changes in spinning conditions. In all such samples examined in this laboratory, an increase in skin thickness is accompanied by an increase in uniaxial orientation, a decrease in selective uniplanar orientation or preferential orientation of the 101 planes, and a decrease in lateral order. This evidence is only indirect but it suggests that the skin is more highly oriented uniaxially and has lower lateral order, while the core has a higher selective uniplanar orientation and lateral order.

Available data support the theory that the skin represents the proportion of the filament which was being coagulated and partly regenerated during the stretch in the region of the spinneret. The core is apparently still fluid viscose during the stretch near the spinneret and is coagulated and regenerated after the skin has set up and has reduced stretchability.

Deswelling and Relaxing

The deswelling or drying process in spinning has a large effect on fine structure. The disordering effect of drying from solvents less polar than water has already been discussed. Water wet cellulose yarns show considerable disordering of the 101 or hydroxyl-rich planes. The effect is observed both in gel cellulose and in rewetted yarn and the measured increase in 101 interference diffuseness appears to be larger than can be accounted for on the basis of mere superimposition of water scattering on the cellulose pattern.

The gel or freshly spun water-swollen yarn can be made either to decrease or increase in orientation on drying depending on the tension applied. Drying under low tensions or in a relaxed condition develops higher lateral order but lower orientation in the yarn than does drying under high tension. Similar changes in lateral order and orientation occur during heating a previously dried yarn in water at 90°C, as noted by Meyer and Mark.²² The change in lateral order takes place during the initial heating in water, either relaxed or under tension. Subsequent wet heating

changes orientation slightly, but has no effect on lateral order. Some very highly stretched yarns show a definite increase in lateral order merely upon wetting and drying relaxed. The changes in physical properties accompanying these changes in lateral order vary depending upon the particular treatment. However, after either drying relaxed or heating in water relaxed the elongation-to-break is increased relative to that of a suitable control. Since both orientations and lateral order change, it is difficult to say just how these latter after-treatment effects fit in with the property-structure correlations already discussed for yarns spun under various conditions. However, as mentioned, it can be said that factors other than orientation and lateral order also have important effects.

The lateral order of freshly spun yarns is therefore metastable in that it is measurably increased by relaxing or heating in water, just as the celluloses of low order obtained by regenerating or deswelling in non-aqueous media were ordered by wetting out at room temperature. In general, however, the disordered structures obtained by a particular mode of spinning have appeared to be more stable than those obtained by non-aqueous regeneration or deswelling. These considerations suggest that lateral disordering introduced after the yarn structure is set up is of a more localized nature than that existing in the original structure.

ACKNOWLEDGMENT

The author wishes to acknowledge with appreciation the helpful suggestions and valuable samples, apparatus, and techniques obtained from Dr. O. E. Schupp and Dr. C. W. J. Wende of the Chemical Department and from members of the Rayon Technical Division, E. I. du Pont de Nemours & Company. Especial thanks are expressed to Dr. N. L. Cox of the Chemical Department for preparing all of the experimental viscose rayons described here, to Dr. A. F. Smith of the Chemical Department for continued guidance and interest in this work, and to the Rayon Technical Division, E. I. du Pont de Nemours & Company, for generous support of this research.

²² K. H. Meyer and H. Mark, *Der Aufbau Der Hochpolymeren Organischen Naturstoffe* (Akad. Verlagsgesellschaft M. B. H., Leipzig, 1930), first edition, p. 183.

A Theoretical Comparison of the Visual, Aural, and Meter Reception of Pulsed Signals in the Presence of Noise*

J. H. VAN VLECK AND DAVID MIDDLETON

Research Laboratory of Physics, Harvard University, Cambridge, Massachusetts

(Received July 9, 1946)

The present paper examines theoretically the relative sensitivity of the detection of signal pulses in the presence of noise, (a) by observation of an oscilloscope, (b) by aural perception, in which one listens to the fundamental or a low harmonic of the pulse repetition frequency (PRF), (c) and by a meter. The metering scheme may be either aperiodic, where the rectified current is fed directly to a meter with a long time-constant, or periodic, where the rectified current is sent through an audio-filter tuned to the PRF, given a supplementary rectification, and then passed through the meter. The dependence of the sensitivities of the different methods on various relevant parameters is studied in some detail. These parameters include the width and the shape of the IF response, the pulse length, the PRF, and in aural or meter reception, the duration of the gate, the width of the audio-filter, and/or the time constant of the meter. The descriptive survey of the results is given in Part I and the mathematical analysis in Part II. Among the more important results are (I). The optimum IF filter is the conjugate of the Fourier transform of the pulse, not merely for visual reception, as was previously known, but also for

aural or meter reception as well. (II): For very weak signals the linear detector requires only about 5 percent more input signal power than does the quadratic to achieve the same minimum detectable signal (same final signal-to-noise ratio). (III): The aperiodic meter has the advantage of not requiring knowledge of the PRF, and has potentially great sensitivity if spurious fluctuations in gain can be balanced out. (IV): Meter methods can be made more sensitive than the oscilloscope if long time-constants are available. Gating is also necessary. (V): Although the best IF filter is the Fourier transform of the pulse, the best pulse is not the Fourier transform of the filter in aural reception (though it is in visual), for the best results in meter or audio-detection are obtained by using long pulses. In visual work, the pulse length is immaterial, to a first approximation. Curves are given showing the power required to achieve a given signal-to-noise ratio as a function of pulse length, IF filter width, the PRF, gating time, and audio-filter width, in some cases when the pulse and filter are not matched (i.e., are not related as Fourier transforms). Some numerical estimates of aural and meter performance relative to visual are also essayed.

PART I. DISCUSSION

I. Introduction and Formulation of the Problem

THE purpose of the present paper is a theoretical comparison of the merits of visual (oscilloscope) and aural detection of pulsed signals, with particular emphasis on the best intermediate frequency (IF) filter design for the two types of reception. It is not *a priori* obvious that the optimum design is the same in the two cases. By visual reception we mean that the modulation is spread out on a screen by means of a cathode-ray oscilloscope. By aural we imply that the emphasis is on detecting a particular tone (the fundamental of the pulse repetition frequency (PRF), or a low harmonic thereof) in the audio-region. Here at the outset we should emphasize that in the present paper we usually employ the term *aural reception* to include any detection device which is actuated by one particular

harmonic component of the PRF, or in other words, involves integration in time rather than the observation of the instantaneous effect of the superposition of all harmonic components, as in visual perception. The instrument need not be the human ear, but instead can be, for instance, a meter, or recording device. The harmonic component, furthermore, under certain circumstances, can be the constant or d.c. term, not involving the pulse repetition period, as in so-called *aperiodic* metering schemes.

In the visual case, the best filter, in the first approximation, is that which gives a maximum value of the ratio of the peak signal to the noise background. This statement can be formulated more precisely as follows: coming into the IF filter there is a fluctuating noise voltage and also a signal $S_1(t)$, which is a definite function of time. After filtering and rectification, the video amplitude is not expressible as the simple sum of pure signal and pure (i.e., normal, random) noise, as there are complicated modulation or interference effects between the two, especially

* This paper is based in whole on work done for the OSRD under contract OEMsr-411 with the President and Fellows of Harvard College.

for the linear detector. If the signal is periodically repeated, there will be a particular epoch at which the amplitude on the oscilloscope screen is on the average a maximum. *The excess of the average of the amplitude at this epoch over the mean amplitude in the absence of signal we denote as s_v .* We say on the average, because sometimes the noise may interfere constructively and sometimes destructively with the signal. We must seek, therefore, to make $s_v(t)$ stand out as much as possible relative to the r.m.s. deflection n_v caused by the noise background. The latter may be computed on the assumption that the signal is zero, since the duty cycle (the "on"-period of the pulse), is presumably very small, and hence the cross-modulation terms contribute little. (See the paragraph following Eq. (16), Part II.) According to this viewpoint the best filter is that which maximizes s_v/n_v . This we shall refer to throughout the paper as the "simple criterion" for visual detection and is that which we shall use as our standard. We employ the adjective "simple" because actually examination merely of the ratio s_v/n_v neglects certain factors. It does not consider the effect of the width rather than the height of a "pip" and of the "blades of grass," and especially it does not adequately take into account the effect of pulse repetition. The more frequent the pulse, the more easily the eye integrates, in the limited time at its disposal, to detect the mean position of the pip relative to a fluctuating background. Integration problems in visual detection are beyond the scope of this work, as they have been considered elsewhere, especially by North,¹ Uhlenbeck,² and others.^{3,4} In discussing the results we shall

¹ D. O. North, in an unpublished report PTR-6C, entitled "Analysis of factors which determine signal-noise discrimination in pulsed carrier systems." (RCA, Princeton, June 25, 1943.)

² G. E. Uhlenbeck and M. C. Wang have carried through similar studies which it is hoped will be available soon at the Massachusetts Institute of Technology's Radiation Laboratory. In this connection, very extensive experimental work has been done on all phases of the visual problem by J. W. Lawson, also at (M.I.T.) Radiation Laboratory. It is expected that this work will be published by 1947, appearing in the appropriate Handbook. (Book 24, Radiation Laboratory Series, "Threshold Signals," by Lawson and Uhlenbeck.)

³ P. J. Suto, "Theoretical effect of integration on visibility of weak signals through noise." (Harvard Radio Research Laboratory, Report 411-77, February 14, 1944.) Available at Harvard War Archives, Littauer Building, Harvard University.

mention the modifications introduced by the more refined criteria.

With audio-reception the point is to make the ratio of energy in the fundamental of the PRF as large as possible relative to the noise in a small frequency interval immediately surrounding the discrete frequency equal to the PRF. The audio-reception may be accomplished in either of two ways. In case the PRF is known, we may insert a filter a few cycles wide immediately surrounding the PRF, so as to cut down the noise. The presence of the signal can then be detected not necessarily by the ear but in principle, say, by lighting a light—more practically, by inserting a thermocouple, or better still, by means of an additional rectification in connection with a meter, as will be discussed more thoroughly in Section IV, Part I. The sensitivity of the device is contingent on the ability to use a very narrow filter, since the signal energy is concentrated at discrete frequencies, whereas noise is spread through a continuum.

In case the PRF is not known, one does not know where to put the filter. The apparatus based on light indication or metering (unless of the aperiodic variety) then obviously becomes quite unworkable in general. There is, however, one instrument which is able to construct a moderately narrow filter about the PRF without knowing where the latter is. This instrument is the human ear, which has the remarkable property that it can pick out energy concentrated in a particular periodic component even though the latter is greatly over-shadowed in total amount by the background energy in the continuum. Namely, the ability of the ear to hear a unifrequent or pure tone f_p is masked only by the noise background in a particular frequency interval $f_p - \Delta f_c/2$ to $f_p + \Delta f_c/2$, centered about f_p , so that effectively the ear acts like a filter of width Δf_c . Of course, the filtering action of the ear is not perfectly rectangular as we assume it, but the error introduced by this idealization is not serious for our purposes. We will not attempt to go into the physiology of why

⁴ E. R. Brill, "Improvement in minimum detectable signal in noise through the use of the long afterglow tube and through photographic integration." (Harvard Radio Research Laboratory, Report 411-84, February 8, 1944.) Available at Harvard War Archives, Littauer Building, Harvard University.

the ear has this behavior, but it is probably due to the fact that it (the ear) embodies a collection of nerves, each responding to a very narrow frequency band, and is capable of discriminating when one nerve is more strongly excited than the average. The effective band width is a function of f_p , the dependence being shown in Fig. 1, taken from data in a paper by Harvey Fletcher.⁶ For the values of f_p in which we are interested, the quantity Δf_s is of the order 50 to 200 cycles. The signals with which one is concerned involve many harmonics as well as the fundamental, but Fig. 1 shows that the effective filter width is greater for higher frequencies, so that they are usually more readily screened by the noise background; hence it is better not to use harmonics, and so we gauge audibility by the masking of the fundamental. It is important to note, however, that when the PRF is much below 1000 c.p.s. it may be preferable to use a suitable harmonic of the fundamental, since the ear is most sensitive to signals in the range 1-2 kilocycles.

According to the above, then, any filter centered around f_p and wider than the corresponding Δf_s is of no particular help, while any filter of width $\Delta f'_s$ narrower than Δf_s has the effect of reducing the power level of the noise by a factor $\Delta f'_s/\Delta f_s$. The point which we wish to emphasize is that *the mathematical criterion for effective reception by the ear is the ratio of signal energy in a discrete Fourier component to the energy in the immediate vicinity of this component*, the "immediate vicinity" being the natural selectivity of the ear or the width of the artificial filter, whichever is narrower. The use of the latter, of course, is possible only if the PRF is known.

In either visual or aural reception there is a certain critical threshold of signal to noise which must be exceeded before the signal is perceived. However, from the above we see that the nature, and so presumably the magnitude, of the threshold is different in the two cases. Namely, for the visual case the critical ratio was that of the average maximum signal to average total video noise amplitude s_s/n_s . On the other hand,

⁶ H. Fletcher, Bell Lab. Monograph B-1205 (Dec. 1938), cf. Fig. 17 of "Auditory Patterns as a Means of Studying the Hearing Process Taking Place when Sound is Sensed."

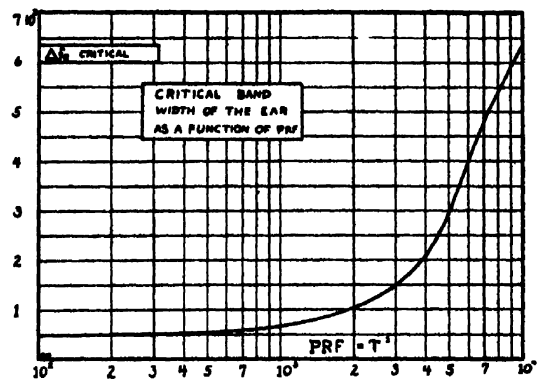


FIG. 1. Critical band width of the ear as a function of the pulse repetition frequency

in aural detection the critical ratio s_a/n_a is that of one harmonic signal component to the r.m.s. amplitude of the portion of the noise which is confined to the spectral region covered by the critical masking zone of the ear (Fig. 1) or by the audio-filter, if narrower.

The question of the numerical results of the minimum detectable values of $(s/n)_v$ and $(s/n)_a$ obviously depends on the observer. In oscilloscope measurements the minimum detectable signal-to-noise ratio is usually expressed by the experimentalists in terms of the amplitude ratio s_F/n_F where s_F and n_F are, respectively, the mean amplitudes of signal and noise after filtering by the IF, but before rectification, rather than in terms of s_s/n_s directly. Very acute observers can detect values of s_F/n_F as low as 10^{-1} , i.e., -10 db, whereas an unskilled person may only see the signal when s_F/n_F is almost unity, which is probably the safer estimate to use under conditions where there are many distractions. (These figures are based on essentially infinite integration time, which for most people is practically of the order of one minute, sometimes running as high as about five minutes under the most favorable conditions before further improvement is unnoticeable.) In the aural case, Fletcher⁶ gives the threshold s_a/n_a as of the order of magnitude unity. Some recent rather crude experiments undertaken by the authors indicate that to perceive a value $s_a/n_a = 1$, the observer must be existing under rather ideal laboratory conditions, and that with any distraction, mental or otherwise, the minimum detectable value of $(s/n)_a$ is more nearly of the order 10^1 (+10.0 db). These

two values, 0 db and +10.0 db (or correspondingly, -10.0 db and 0 db in the visual case), we shall refer to as the "optimist's" and "pessimist's" estimates, respectively.

The criteria of the different minimum detectable signals for visual and aural reception may now be specified in analytic form. Throughout the paper we are unconcerned with the first rectification (or mixing) as this is essentially a linear process which has the effect of displacing the carrier frequency from an RF value f_r to an IF value f_0 . After the IF filter the mixture of signal and noise is sent through a rectifier, here taken alternatively as quadratic (characteristic, $f(x) = x^2$), or linear, without bias (with the characteristic $f(x) = x$, $x > 0$, $f(x) = 0$, $x < 0$). The signal disturbance is denoted by $s_v(t)$. In visual detection the minimum observable signal-to-noise ratio is some particular value of $s_v(t)/n_v$, where n_v^2 is the total video noise power, given by

$$n_v^2 = \int_0^a W_{LF}(f) \frac{d\omega}{2\pi},$$

where a is a limit great compared to the IF filter breadth but small compared to $2\pi f_0$, and where $W_{LF}(f)$ is the spectral density of the mean square noise amplitude. (See Eqs. (18), (26), Part II.)

In audio-detection, on the other hand, one is concerned with the energy in a given harmonic component, divided by the energy in the immediate region surrounding this component. Therefore the effective signal-to-noise ratio is

$$s_a/n_a = \frac{2^{\frac{1}{2}}}{T} \left| \int_{-T/2}^{T/2} s_v(t) \exp(ik\omega_p t) dt \right| / [W_{LF}(kf_p)\Delta f_s]^{\frac{1}{2}},$$

where k is the order of the harmonic component kf_p being received, $\omega_p = 2\pi/T$ is the pulse repetition (angular) frequency, and Δf_s is the filter width. (The filter is regarded as rectangular; if it is not, an equivalent rectangular width which admits the same amount of power for a filter which has the same maximum response is to be used.)

Throughout the paper certain simplifications are made in the expression for s_a/n_a . In the first place, it is assumed that the pulses do not over-

lap appreciably, which is certainly true if the width of the IF filter is large compared to the PRF. The limits of integration in taking the Fourier component can then be replaced by $+\infty$ and $-\infty$, appropriately, if $s_v(t)$ is then understood to be only the disturbance due to one pulse. Further, the length (in time) of one pulse is supposed short compared to the pulse repetition period, so that the factor $\exp(ik\omega_p t)$ may be taken as nearly constant in the region where $s_v(t)$ is appreciably different from zero. This is true unless the harmonic is so high as to be comparable with the reciprocal of the pulse length. Since the PRF is small compared to the noise band width, we may assume that the noise background $W_{LF}(k\omega_p)$ has substantially the same value as in the vicinity of zero frequency, making the above expression for s_a/n_a take the form

$$s_a/n_a = \frac{2^{\frac{1}{2}}}{T} \int_{-\infty}^{\infty} s_v(t) dt / [W_{LF}(0)\Delta f_s]^{\frac{1}{2}}.$$

This expression is independent of k ($k = 1, 2, \dots$), so that it is immaterial which harmonic component is utilized.* The input power required to achieve a given critical value of the ratio $(s/n)_c$, or $(s/n)_a$ will be a function of the following variables:

- the width and shape of the IF,
- the width and shape of the pulse,
- the pulse repetition frequency (PRF),
- the duration of the gate,
- whether the rectifier is linear or quadratic,
- whether the detection is visual, aural, or by meter, and
- the width of the audio- (or meter) filter.

It is the purpose of this paper to examine how the input power depends on any one of the above variables, when the other six are held fast.

II. Best IF Filter Width

We begin by studying the optimum design of the IF filter. Here there is a very simple result. Namely, the best IF filter is one which is the conjugate of the Fourier transform of the pulse

* Here the choice of harmonics is unimportant from the point of view of the *intensity* level of the signal but not from the *loudness* level, which is a measure of the actual perception of the ear, and which varies with frequency. The intensity level merely measures the sound energy per unit area falling on an intensity-level meter, and is not the subjective intensity, or "loudness" of the perceived sound in the ear.

shape (except possibly for a phase shift which varies linearly with frequency). That is to say, if the modulated pulse presented to the IF is $\sqrt{2}s_I(t) \cos \omega_0 t$, and if the Fourier transform of $s_I(t)/\sqrt{2}$ is denoted by

$$S_I(\omega) = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} s_I(t) e^{-i\omega t} dt,$$

so that

$$s_I(t) = \frac{1}{\pi\sqrt{2}} \int_{-\infty}^{\infty} S_I(\omega) e^{i\omega t} d\omega,$$

then the best filter design is one whose amplitude response is

$$G_{IF}(\omega)_{OPT} = \alpha S_I(\omega)^* \exp(-i\omega t_0).$$

Here α is a trivial constant factor calibrating the absolute sensitivity of the filter. The origin of frequency is taken at the carrier, so that the point of maximum response is $\omega=0$. Usually the pulse is symmetrical about its maximum, and then G_{IF} is real, the distinction between $G_{IF}(\omega)$ and its conjugate $G_{IF}(\omega)^*$ vanishing. This result is demonstrated in Part II, Sec. III, cf. Eq. (53).

The criterion is general. It applies regardless of whether one has audio- or video-reception, or whether one has a linear or square-law detector. The existence of the Fourier transform criterion has already been noted, in the visual case by North¹ and others. In the instance of aural reception, the phase of $G_{IF}(\omega)_{OPT}$ is immaterial, and it is only necessary that $|G_{IF}(\omega)_{OPT}| \sim |S_I(\omega)|$ (vide Eq. (56)).

The phase shift constant t_0 is arbitrary as far as the optimum quality of the filter is concerned, i.e., it does not affect its performance. A change in the value of t_0 merely gives a different time delay of a pulse transmitted by the filter. Actually, any physical filter will have a definite value of t_0 , as discussed in more detail in Sec. III of Part II, but knowledge of t_0 has no bearing on the discussion of filter efficiency.

It is particularly interesting that the Fourier transform criterion applies even if the reception is by audio means. The aural case does not appear to have been too clearly understood. Some, for instance, have thought that it would be best to have a very narrow filter (much narrower than the Fourier transform of the pulse) just because in aural reception one is

immediately concerned with only a narrow band of frequencies. However, this is not correct—the best IF filter will be just as wide as for video reception. See Sec. III (b), Part II, for the details of the demonstration.

A filter whose characteristic satisfies the optimum condition, i.e., is the conjugate of the Fourier transform of the pulse, we shall henceforth refer to as a *matched* filter. It has sometimes been suggested that in visual reception results will be improved by adding a video-filter, i.e., a filter following rectification. If, however, the IF filter is already matched, the insertion of such a video-filter theoretically gives no improvement. (See Book 24, M.I.T. Radiation Laboratory Series, when it appears.)

III. The Linear vs. the Quadratic Detector

On the whole the results for the half-wave linear and quadratic rectifiers are remarkably similar either in aural or visual reception. This is particularly true if one has sensitive apparatus and/or good observers, so that the values of the signal-to-noise ratio s_F/n_F are reasonably small compared to unity. Then the behavior of the linear and quadratic devices is the same as regards the value of s_F/n_F , corresponding to a given pulse system and IF filter, except for a numerical factor. The latter proves to have a value $0.957 = \frac{1}{2}(4/\pi - 1)^{-1}$, as has already been noted by North¹ in his interesting paper. In the audio case the constant is even more nearly unity, being somewhat dependent on pulse shape. For the specific example of gaussian pulses and filters, considered analytically in Secs. IV and VI of Part II, the constant is 0.973. The result in the video cases is contingent only on small s_F/n_F and does not demand a particular filter shape, (e.g., gaussian), or matching of the IF filter to the pulse. The mathematical reason for this similarity in the video case is that for small values of s_F/n_F the signal amplitude transmitted by the rectifier is given by a monotonically increasing series in the ratio s_F/n_F of the form $0.957[(s_F/n_F)^2 + \dots]$. If the first term only is kept, a quadratic behavior follows, like that of the square detector whose first term is $(s_F/n_F)^2$, Eq. (46). The absence of the linear term s_F/n_F is the analytical embodiment of what the engineer

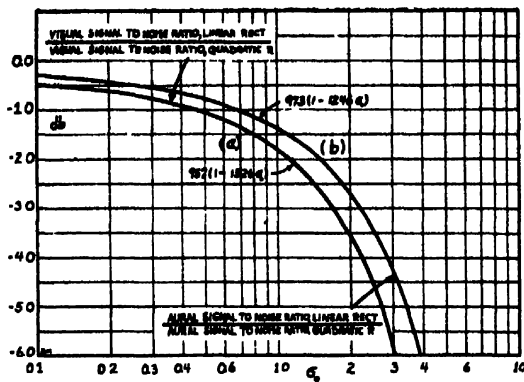


FIG. 2. A comparison of the linear and quadratic detectors in aural and visual reception at match. Curve (a) is a comparison of the linear with the quadratic detector for visual reception, while curve (b) is the same comparison in aural reception, a gaussian pulse being assumed. Here σ_0 is the ratio of the peak signal to r.m.s. noise amplitude leaving the IF.

calls *modulation suppression*, and makes the transmitted signal very much less than it would be otherwise. (See Sec. II (d), Part II.)

The retention of only this leading term, however, always overestimates the performance of the linear detector. A more accurate expression for not too large values of s_F/n_F requires that the inferiority factor for the linear rectifier in video reception be $0.957(1 - 0.1526s_F/n_F)$ instead of 0.957. This is a sort of interpolation formula valid up to about $s_F/n_F = 10^1$. For large values of s_F/n_F the signal amplitude becomes obviously linear in s_F , and the noise rather than the signal is suppressed. The second term of the interpolation formula represents, so to speak, a correction for the tendency of the instrument's response to become ultimately linear rather than quadratic. The relative performance of the two types of detector in the video case is shown in curve (a) of Fig. 2, which is obtained by dividing Eq. (47) by (46) (see also Eq. (82)).

In the aural case a general expression for the relative efficiency of the linear and quadratic detectors cannot be given except for small s_F/n_F until the shape of the pulse and filter are specialized. One can, however, say that when s_F/n_F is not appreciable, i.e., $s_F/n_F < 1.0$ the inferiority factor of the linear detector is always less than about 1.5 db, and the deviations from quadratic performance are of the same order of magnitude as in the video case. This is discussed at some-

what greater length in Part II, Secs. II (e), and Sec. IV (c). Curve (b), Fig. 2 is obtained from Eq. (84) and illustrates the relative aural performance of the detectors, when the incoming pulse is gaussian.

IV. Use of a Meter

It is particularly to be noted that with aural reception the background noise power n_a^2 is directly proportional to the width of the audio-filter. Hence, in principle, if one can construct arbitrarily narrow filters and if the signal is exactly periodic so that its Fourier component is strictly a "delta"-function (i.e., a function with an infinitely high and narrow peak at some point $x = x_0$, and zero everywhere else, $x \neq x_0$) one should, for a given minimum detectable value of $(s/n)_a$, be able to achieve infinitely high sensitivity in terms of s_F/n_F , simply by narrowing the audio-filter indefinitely. However, there are of course, practical limitations as to how much a filter can be narrowed, viz. (a), purely constructional difficulties, (b) insufficient stabilization of the PRF, and (c) limitations because the targets are not in the path of the beam an infinitely long time, so that we have a signal repeated a finite rather than an infinite number of times. The signal spectrum consequently is not uni-frequent, but is "smeared out," having Fourier components which are distributed over a finite interval instead of being sharp delta-functions. Probably (b) is not too serious a difficulty. As regards (a) it is not easy to build audio-filters even as narrow as 10 cycles, and still narrower filters would require lock-in amplifiers, which are fairly clumsy and bulky pieces of apparatus. Effect (c) is always present to some extent.

Aperiodic Methods

Although a narrow filter is difficult to construct in the audio-region, on the other hand, a very narrow low pass filter, which includes d.c., is easy to obtain simply by sending the rectified current through a condenser in parallel with a meter, which is used as the recording instrument. Consequently, the idea may occur to detect the constant term of the Fourier expansion of the signal, rather than the fundamental. As the constant term does not involve the PRF in any way,

instruments based on the recording of this term may be called "aperiodic."

The constant term in the Fourier expansion has associated with it an energy half as great as any other component.* This reduction by a factor $\frac{1}{2}$ is offset by the fact that the low pass filter extends on only one side of the origin in frequency, so that the noise continuum transmitted is halved. Hence our various expressions for signal-to-noise ratio, etc., in aural detection also apply to the aperiodic metering scheme if we use in the formulas an effective filter width equal to *double* the width of the physical low pass filter.

As far as the competition between the signal and the fluctuations in the continuous noise background is concerned, the zero frequency (also called aperiodic) method is thus satisfactory. The difficulty comes from masking due to spurious variations in the d.c. due to noise, or in other words, from what the engineer calls "changes in gain." When we say spurious, we mean variations caused by factors (e.g., temperature, apparatus changes, etc.) other than statistical fluctuations like those in the Shot effect, etc. From the purely statistical standpoint, the d.c. due to noise is constant in time, and can, in principle, be balanced out by a proper bias. However, the d.c. due to the noise is enormously larger than the continuous noise background admitted by the filter, and hence much greater than the signal, as the latter two quantities are comparable. Namely, the energy in the d.c. exactly equals that in the *whole* low frequency noise continuum, for a square detector and exceeds it by a factor $3.7 = (4/\pi - 1)^{-1}$ for the linear rectifier. Since the width of the low frequency continuum is of the order of the IF frequency width we see that the d.c. noise will exceed the fluctuating noise continuum admitted by the filter by a factor of the order of the ratio of the IF filter width to that of the audio. This ratio is about 10^4 – 10^6 . Hence an infinitesimal spurious variation in the d.c. noise voltage will overshadow and obscure the modulation due to

the signal, and so make the device not at all feasible, unless it has very special equipment for balancing out spurious fluctuations.

If the oscillations in gain are very slow, they can be avoided simply by inserting a blocking condenser. The filter is then effectively not of the low pass type, but instead cuts off at very low frequencies, so that pure d.c. is eliminated. This artifice will work only if the spurious changes in gain have a longer "time constant" than does the signal modulation, so that the filter will pass the latter but not the former. Unfortunately, the two times are rather likely to be comparable in many applications.

Periodic Methods

There is one way of escaping from this difficulty of gain fluctuations, and this is to use a meter in conjunction with an audio-filter. This procedure requires rectification. Current confined to the frequency range passed by the audio-filter (centered about, say, the fundamental of the PRF), is rectified, and enters an electromechanical low pass filter which admits only d.c. and extremely low frequencies (for example less than $\frac{1}{10}$ or 1 cycle as an upper limit). To avoid constructional difficulties, the audio-filter will not be nearly so narrow as the meter filter. The presence of the signal then shows up as a change in the amount of d.c. It may seem at first sight that this scheme is exactly like the aperiodic one just discussed, but the insertion of the additional rectification* makes a great difference. In other words, the action of

(A) rectifier→low pass (directly) is quite different from

(B) rectifier→audio-pass→rectifier→low pass.

Obviously (B) requires knowledge of the PRF, whereas (A) does not. Hence (A) and (B) may be called, respectively, the aperiodic and the periodic metering methods.

In (A) as in (B) after the final rectification, the energy in the d.c. due to noise is about equal to the energy in the continuum. The continuum, however, has a width comparable in (A) with

* The reason is that the constant term has an amplitude $(1/T) \int_{-\infty}^{\infty} s_c(t) dt$, rather than $(2/T) \int_{-\infty}^{\infty} s_c(t) dt$, but the reduction in intensity is by a factor $\frac{1}{2}$ instead of $(\frac{1}{2})^2$, because for the constant term there is no distinction between r.m.s. and peak amplitude.

* Note: The audio-filter is always sufficiently narrow that the noise entering the final rectifier may be considered "pure" and random, for if a small portion of the spectrum alone is taken, randomness is restored, and the gaussian characteristics of pure noise are manifest.

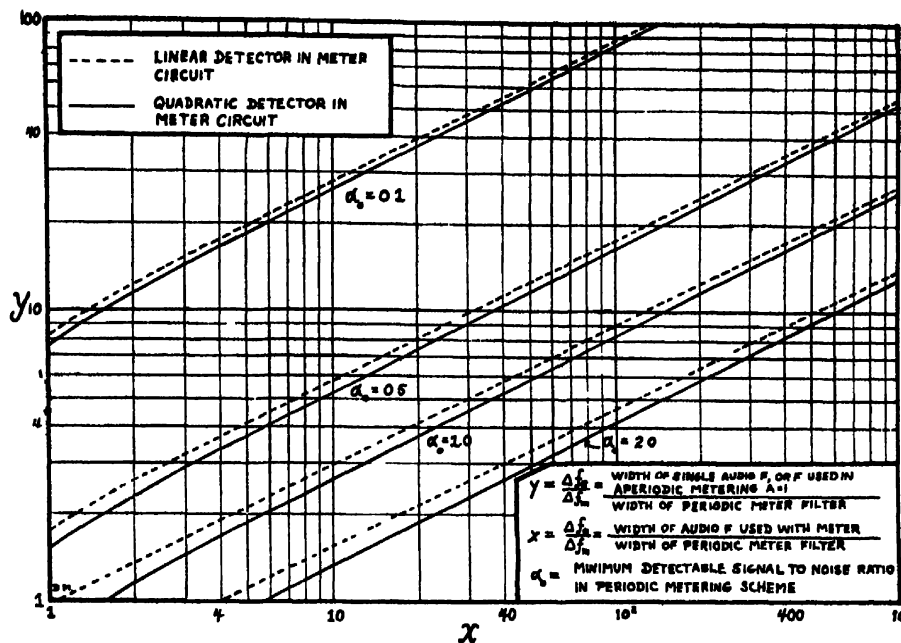


FIG. 3. Width of the equivalent earphone filter Δf_s used in aural detection of the minimum perceivable signal in the presence of noise as a function of the audio-filter width Δf_m used in meter detection of the same signal-to-noise ratio α_0 . The two widths compared are both normalized in the figure by division by the quantity Δf_m , which is twice the width of the final low pass filter in the periodic metering scheme. It is assumed that the minimum discernible signal-to-noise ratio in the final circuit is the same for aural or meter detection.

the width of the IF filter, and in (B) with that of the audio-filter. Hence the final continuum is much narrower in (B) than in (A). As a result, the low pass filter admits a very small fraction of the total continuum energy in (A) but an appreciable fraction of it in (B). Assuming that the minimum detectable signal (apart from troubles due to spurious variations in d.c.) corresponds to the low pass filter receiving a signal comparable with the noise background, we see that in (B) the signal will be comparable (say within a factor 10 or 100) with the d.c. of the noise, instead of being a million or so times smaller than the latter. Thus spurious variations in d.c. intensity will not be lethal. (We have here assumed that the discrete noise at the PRF in the audio-filter has been balanced out by some method such as the Jordan double-gate scheme, described later. If this is not so, the signal may be more nearly one-thousandth or one ten-thousandth of the noise d.c., but still the latter is not so overwhelming as in the straight aperiodic metering system.)

The mathematical theory of the extra rectifica-

tion involved in the periodic meter is treated in Sections V and VI, Part II. The theory is in the main the same as that for the rectification following the IF, with the audio-filter now playing a role similar to that which the IF filter did previously. However, there is one simplification. The signal transmitted by the audio-filter is sinusoidal rather than pulsed, so that its amplitude can be treated as constant.

There are some false impressions to be guarded against in considering the periodic meter. In the first place, for the periodic system (B), a very narrow reception band in the final circuit (e.g., 0.05–0.5 cycle), with a narrow audio before the extra rectification (e.g., 5–10 cycles), does not represent nearly so great a sensitivity as would be obtained if one had such good filter construction that one could put an equivalent very narrow band (say, 0.05–0.5 cycle) directly in the audio-circuit. The reason is the non-linearity of the rectification. Namely, if the audio signal-to-noise ratio is small (and if it were not, there would be no need for the meter), the response of the rectifier is essentially quadratic in this

ratio, even if it is of the linear rather than the square variety, because of modulation suppression, as shown by Eq. (97).

Curves illustrating the effect discussed in the preceding paragraph are given in Fig. 3. They are drawn for a linear rectifier in the meter circuit (dotted line), but the results are not materially different for the quadratic case (solid line), especially for weak signals. These curves correspond to various assumed minimum detectable signal-to-noise ratios for the meter, and the abscissa is the ratio of the audio to meter filter width. The ordinate is the audio-filter width, measured in multiples of the meter filter width, which would give the same signal-to-noise ratio in the audio-circuit, without the meter, as obtains with the meter in the final circuit. In other words, the ratio of ordinate to abscissa measures the factor by which the audio-filter would have to be diminished in order to permit one to dispense with the meter circuit. For the typical case of a minimum detectable signal-to-noise of unity, a linear second detector, and a meter width of one cycle, $\Delta f_m = 2$, and with a ten-cycle audio-filter $\Delta f_a = 10$ the equivalent single audio-filter is 4.3 c.p.s. As an example of another typical case, assume again that the minimum detectable signal-to-noise ratio is unity, the width ($\Delta f_m = 1/\tau_0$) of the meter filter is 1/20 cycle, and that of the audio is now 5 cycles. The equivalent single audio-filter is then 0.65 c.p.s. and the equivalent low pass filter is 0.33 c.p.s. These figures are based on the assumption that the filters are of gaussian shape, and the width is determined by the half-power points. In case all the filters are nearly square, or "ideal," the formulas still apply if one takes the mathematical quantity Δf to be 1.06 times the actual filter width.

Considerations Relative to Time-Constants of Meter

In either periodic or aperiodic methods the noise is counterbalanced by a suitable bias, and the *mean* meter reading can thus be made zero in the absence of signal. The presence of the signal will cause a systematic deflection of the needle, if the time-constant of the meter is small. Thus for a comparatively wide meter filter, the

limiting factor preventing observation is the irregular fluctuation due to the noise continuum.*

The question of how much trouble is caused by the long time-constant of a meter circuit appears to be appraised very differently with a discouraging disparity of opinion. Some claim extraordinary sensitivity with meter reception, whereas others claim that the time-constant difficulties are so serious that, for practical purposes, a meter circuit is to all intents and purposes useless on an airplane or for other operations. Perhaps a conservative viewpoint is that a meter circuit will not give too much gain in efficiency unless the target is under observation in the receiver for at least a second and this, of course, requires that the angular scanning be rather slow. But in circumstances where long time-constants, say one to six or eight seconds or more, are allowable, the (periodic or aperiodic) meter, or its cousin, the recording tape, may be able to compete favorably with other methods of perception, for example, with aural detection where the strain of "listening through" is much greater than for "looking through." There seems to be insufficient experimental evidence at this time, however, as to precisely what a meter can and cannot do compared with that for other systems. So before a definite answer can be given, we must know at least what the order of magnitude of the minimum perceivable signal on the meter is. At any rate, one is certainly not *a priori* justified in discarding meters as detecting devices.

The experimental evidence available to us at present seems to indicate that meters with short time constants, i.e., wide filters, say, 1-2 cycles, are not as effective as other methods when the situation is such that the time of observation of a

* Note: One point should be noted in connection with these fluctuations. They are themselves modified by the presence of the signal, as the cross-terms between signal and noise make the continuum energy different from what it is with noise alone. This effect was wanting in previous parts of the analysis, as the pulsing of the signal made the signal-noise cross terms unimportant because they existed only a small fraction of the time. The meter, however, is exposed solely to a sinusoidal signal since the audio-filter excludes the harmonics which interfere with the fundamental to shape the pulses. We do not endeavor to calculate the modification of the continuum energy by the signal (which one of us hopes to treat in a later paper), because it is difficult, because experimentally it does not seem to be an effect of much importance, and because it is not clear whether the increase of the "wobble" (in distinction from systematic drift) of the indicator when the signal is on makes the latter easier or more difficult to perceive.

target is brief and the conditions difficult. However, in operations where longer intervals of observation are allowed, say 2-10 seconds, or longer, corresponding approximately to meter filter widths of $\frac{1}{2}$ to $\frac{1}{10}$ cycles the meter may offer definite advantages over other systems of perception. For an example, consult the table, Sec. VII, Part I.

Clearly, the minimum detectable signal-to-noise ratio in a meter depends a great deal on tactical considerations and will be higher in the field than in the laboratory. In the table, Sec. VII, we have assumed that s_m/n_m is unity. This whole question of the minimum perceivable signal in meter detection, as well as in the other systems, devolves ultimately upon memory, whether it be psychological, i.e., remembering a series of successive readings, or mechanical, in which the memory process is achieved by means of a recording device, such as an Esterline-Angus recording meter, for example. Under ordinary conditions a meter signal-to-noise ratio s_m/n_m of unity seems a reasonable guess. However, in cases where the memory of the observer is particularly good or when, say, recording tapes are used, the value of s_m/n_m may be considerably less than 1. In other words, high sensitivity may also be secured by waiting for a sufficiently long period of time, even though the time-constant of the meter itself is not unduly long.

Use of a Gate

In order to build a very sensitive aural or metering scheme, one should not only have a very narrow filter or long time-constant, but in addition one should employ a gate. The effect of the gating device is to reduce the noise background by the "gating factor," i.e., the fraction of the time that the gate is open or transmitting. The gate, however, does not reduce the noise while the signal is being received. Hence the gating factor should not be used in computing the amount of modulation suppression of the signal in a linear rectifier due to the simultaneous presence of noise. When the gate has the same "on-off" period as the signal, it creates a periodic disturbance with the same period as the signal, and so might easily be confused with the latter in aural reception or in detection by a periodic metering scheme. However, there is the ingenious

idea, due to Dr. W. H. Jordan (M.I.T. Radiation Laboratory), of opening the gate twice as often as the signal is repeated. The periodicity in the noise, arising from the gating, then has a fundamental equal to the second harmonic $2f_p$ of the signal period f_p , and so will not be heard at all when we listen to the fundamental of the PRF.

With ordinary video or cathode-ray presentation a gate is of no appreciable help as it does not affect the signal-to-noise ratio *while* the signal is being displayed. On the other hand, audio-detection examines a particular Fourier component, and compares the latter with the average noise background rather than just that in the presence of the signal. The insertion of a gate is here a very material help. The ability to obtain as great a sensitivity with the audio as with the video method is usually contingent upon the ability to use a rather narrow gate in the former, corresponding to a gating factor of the order 10^{-2} or so.

V. The Dependency of Visual and Aural Response on Filter Width and Pulse Duration Away from Match

We have mentioned in Section II that the result that optimum performance is obtained at match (i.e., when the filter response is the conjugate of the Fourier transform of the pulse) is independent of the spectral "shape" of the filter. However, when we wish to compare the performance at mismatch with that at match, the behavior depends somewhat on the shape of the pulse or the filter. In order to make explicit calculations possible, we henceforth assume throughout Section V (Part I), that the filter and pulse both have a gaussian structure* so that their amplitude-frequency characteristics are, except for constant factors, $\exp(-\omega^2/2\omega_b^2)$ and $\exp(-\omega^2/2\omega_p^2)$ respectively (where the origin of frequency ω is at the carrier). As the Fourier transform of a gaussian function is also gaussian, the pulse depends on time in a gaussian fashion, *vis.*, $\exp(-\omega_p^2 t^2/2)$. At match one has $\omega_b = \omega_p$, and we now investigate how much the performance is jeopardized by having mismatch, where $\omega_b \neq \omega_p$.

* Note: Our choice of the gaussian response is based on the observation that such a characteristic is a good approximation to actual ones in practice, except for frequencies well away from resonance. (See Sec. III, Part II.)

Let us imagine first that the pulse width is given and the IF filter width is varied, so that ω_p is fixed and ω_b is changed. The power necessary to achieve a given minimum detectable signal is readily calculated and may be expressed simply in terms of the ratio of power necessary at mismatch to that required at match. The details of the analysis have already been carried through independently by Wang and Uhlenbeck at (M.I.T.) Radiation Laboratory, so we will omit them in this paper.

Figure 4 shows the necessary power as a function of $\lambda (=f_b/f_p)$ for the case of aural reception. The curves for the linear detector are not the same as for the quadratic rectifier; namely, the shape of the curve in the linear case is a function of the signal-to-noise ratio $\sigma_0 = (s_F/n_F)$ at match, which in turn depends on the minimum detectable signal. The dependence on σ_0 in the linear detector occurs because the value of σ_0 determines the amount of modulation suppression. If, however, the signal is very weak, the performance of the linear rectifier is the same as that of the quadratic except for a factor independent of λ . Comparison of the curves in Fig. 4 with the results for the visual case indicates that aural reception is somewhat less sensitive to mismatch.

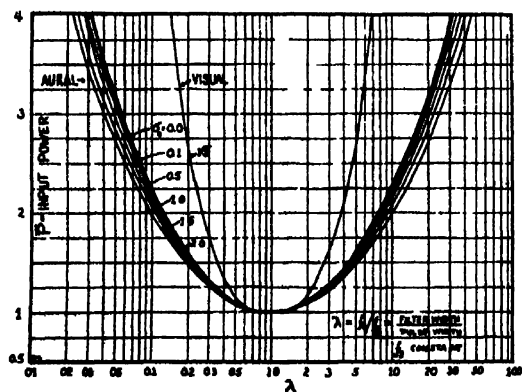


FIG. 4. Input power to the IF as a function of IF filter width at constant pulse width for aural reception of the minimum perceptible signal (Eq. 78b), with a linear rectifier. P is defined as the ratio of input power to input power at match. A gaussian pulse is assumed. The curve labelled $\sigma_0=0$ also applies for all values of σ_0 in the instance of a quadratic rectifier (Eq. 77b). A curve for visual reception is also included.

Let us now see what happens when the filter width is fixed and the pulse width is varied. In the case of visual reception the calculations show that the performance involves the filter width Δf_b and pulse length $\Delta T_p = 1/\Delta f_p$ only through the product $\Delta f_b \Delta T_p$. Hence in the visual case it makes no difference whether we alter filter width or pulse length;* more power is always required when a condition of mismatch exists.

On the other hand, in aural reception the curve obtained by varying pulse length is of a quite different shape from that obtained by varying the filter width. The reason for this is that the quantities Δf_b and ΔT_p here do not enter purely in the product $\Delta f_b \Delta T_p$. Our proof of maximum efficiency at match applies only when Δf_b rather than ΔT_p is varied. If we define $\bar{\lambda}$ as the ratio $\Delta f_p/\Delta f_b$ of the frequency width of the pulse to its width at match, or, in other words, as the reciprocal of the ratio of the pulse length to its length at match, then the variation of necessary power with $\bar{\lambda}$ is shown in Fig. 5 for the linear and square detectors. The curves do not have minima at $\bar{\lambda}=1$. The necessary power is always diminished by lengthening the pulse. However, for extremely long pulses the necessary power approaches an asymptotic value which Fig. 5 shows is nearly achieved when $\bar{\lambda}$ is less than 0.1 or so. Here again the results for the linear detector depend on the signal-to-noise ratio σ_0 , with the curve for $\sigma_0=0$ the same as for the quadratic rectifier. The upshot of Fig. 5 is that *long pulses are more easily heard, or detected by a meter than are short ones.*† Of course there are practical

* Within the approximations assumed in our criterion; see section 1 of part I.

† Possibly the reader is concerned with the physical reason why the dependence on pulse length is so different for the visual and aural cases. The explanation runs as follows: It is clear that in either case the power requirements will be increased if the pulse is made too short, for with very short pulses the Fourier spectrum becomes so wide that only a fraction of the energy is passed by the filter, with attendant wastage. In the visual case it is also wasteful to have the pulse too long, for with our "simple criterion," all that counts is the peak value of the pulse after filtering. Keeping the pulse on a very long time, i.e., much greater than the reciprocal of the frequency band width of the filter, will not appreciably augment this peak, but will increase the energy that has to be supplied, just because of the greater time duration. Hence with visual detection the pulse should be neither too long nor too short, and there is thus some particular pulse length at which the input power required to achieve a given signal-

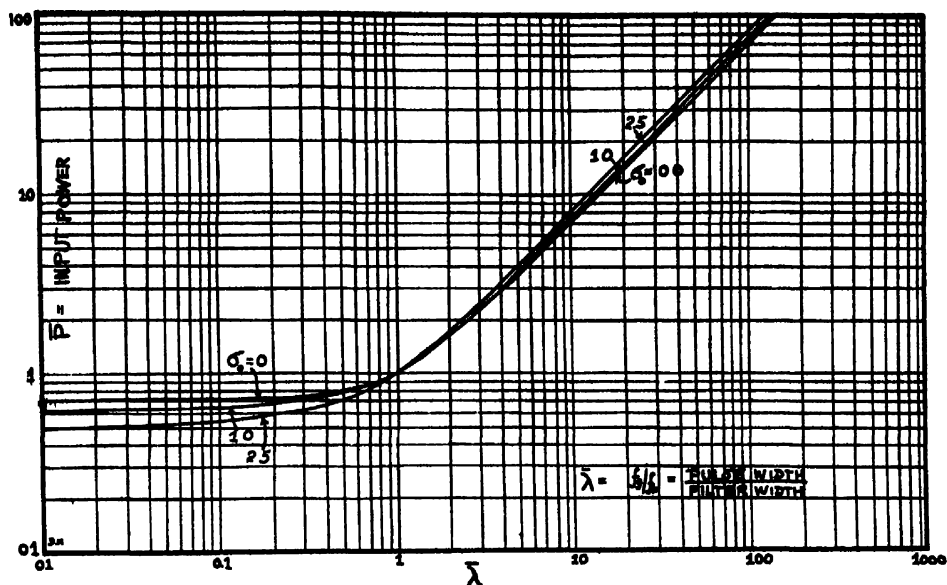


FIG. 5. Input power to the IF as a function of pulse width at constant filter width for aural reception of the minimum perceptible signal, Eq. (78a), with a linear rectifier and gated or ungated noise. The pulse has a gaussian shape. The width in frequency is here the reciprocal of the duration of the pulse. Again the curve $\sigma_0=0$ applies for all values of σ_0 in the case of a quadratic rectifier, Eq. (77a).

limits as to how much the pulse length can be increased, for if the pulse length is too great, there will be an overlapping of successive pulses, and the received intelligence would then be destroyed, not to mention the fact that mathematically our calculations are all predicated on the assumption of no overlapping.

to-noise ratio is a minimum. In aural detection we are interested in the energy in one particular Fourier component. The total number of Fourier components of appreciable magnitude is of the order of the ratio of pulse period to pulse duration. Hence, if the pulse is lengthened the energy is distributed among fewer components, or, in other words, there is less wastage in the high harmonics, which are of no concern when we listen to the fundamental or a low harmonic. This effect counterbalances the fact that the greater duration of long pulses demands more power. Consequently, the power requirements approach asymptotically a limiting value when the pulse is very long, rather than increasing as in the visual case.

If, instead, the pulse length is kept constant and the filter is varied in width, it is clear that too wide a filter is wasteful, as it simply increases the total noise without improving the transmitted signal. At first thought it might seem that narrowing the filter a great deal would diminish signal and noise equally. However, the coherence of the various Fourier components of the signal (before rectification) as compared to the incoherence of the noise has the result that the rectified signal is weakened more than the noise by the narrowing, so that for an infinitely narrow filter the signal-to-noise ratio becomes zero, and consequently in Fig. 4 the power requirements become infinite as λ approaches zero.

VI. The Influence of PRF, Gating, and Audio-Filter Width on Sensitivity

Visual Reception

The dependence of the input power on PRF in visual reception can be dismissed rather summarily. Namely, with our "simple" criterion as to the minimum detectable signal, the only thing that counts is the energy per pulse. The mean power requirements are therefore directly proportional to the PRF, while the peak power will be independent of it. This is what one would expect, for with our "simple" criterion for measuring visual detection, which neglects the integrating effect of the eye (and brain), the only thing that is important is the instantaneous ratio of maximum transmitted signal energy to the noise background, and it is immaterial how many pulses are generated per second. Hence it is most economical to have the pulses spaced at very long intervals.

Neglect of the effect of integration in the visual case will have no bearing on the relative merits of the linear and square detector, and will not do much harm in considering questions of filter structure. On the other hand, in considering the

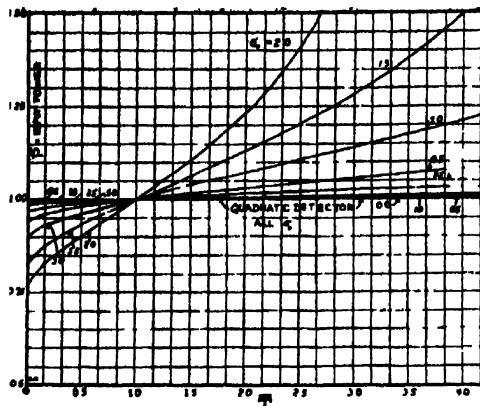


FIG. 6. Input power to the IF as a function of pulse period ($T = T/T_0$) for aural reception at match, of the minimum perceptible signal, with *ungated* noise and linear rectifier, Eqs. (77a, 77b). Here T is the ratio of the pulse period to the pulse period at match.

dependence of visual efficiency on PRF it is certainly important to include the integration effects which we neglect. The theory of North and of Uhlenbeck indicates that at least over a certain range of PRF the visibility of a pulse is proportional to the square-root of the PRF. This has been borne out by the extensive experiments of Lawson (M.I.T. Radiation Laboratory). Hence the effect of integration is to make the average power necessary to achieve a minimum detectable signal proportional to the square root rather than to the first power of the PRF, and to make the peak power inversely proportional to the square root of the PRF.

Audio-Detection—No Gate

In the audio case the precise formulation of the minimum detectable signal presents itself in a more clear-cut fashion inasmuch as our criterion in terms of a particular Fourier component already includes the effect of integration. There still is, to be sure, a question as to what is a proper numerical value of the critical threshold.

Because the gating factor $G = \tau_0/T$ is directly proportional to the PRF, we must be careful to distinguish between those cases for which gating is or is not present in aural presentation.

First, let us assume that there is no gate. With no gate, theory shows (see Eqs. (67), (68)) that for the square-law device, the minimum detectable signal is independent of the PRF and at

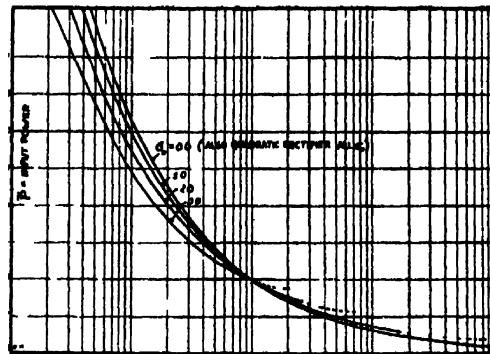


FIG. 7. Input power to the IF as a function of the pulse period $T = T/T_0$ for aural reception of the minimum perceptible signal, Eqs. (77a, b), at match with *gated* noise and a linear rectifier. The curve for $\sigma_0 = 0$ also applies for all σ_0 in the case of a quadratic rectifier, Eq. (80).

match is a function only of the input power to the IF and of the IF filter width. This behavior can be explained in the following way. For a given energy per pulse and a given IF width, the longer the interval between pulses the less power consumed, but the power is less effective because the Fourier components are more closely spaced, and hence there is less energy in any one component. More precise analysis shows that with a quadratic rectifier these two effects just cancel, so that for a *given pulse duration* and IF filter width, the input power to the IF required to achieve a particular value of the minimum detectable signal-to-noise ratio s_n/n_a is not affected by the PRF.* This conclusion holds quite irrespective of whether the filter is matched or the pulse is gaussian in shape. For the linear rectifier this independence of PRF holds only in the limiting case of a very small value of s_F/n_F . Actually, for non-vanishing s_F/n_F the performance with a linear detector is a slowly varying function of the pulse repetition frequency, the input signal being more easily detectable when the PRF is increased as is shown in Fig. 6. The curves of Fig. 6 are drawn on the assumption of a gaussian pulse with matched filter, but the trend with PRF will also be the same in other cases, as the exact filter structure merely determines the precise degree of modulation suppression, and so is reflected mainly in the absolute power scale.

* We are speaking here and elsewhere of mean power. The peak power is correspondingly inversely proportional to PRF.

Insertion of a Gate

The conclusions of the preceding paragraph also apply as long as the PRF is varied but the gating factor (percentage of time the gate is open) is held constant, as then the gate merely changes the absolute sensitivity of the instrument. Physically, however, one usually varies the PRF, keeping constant the absolute time the gate is open, rather than percentage time. Then the dependence on the PRF is more complicated and needs discussion.

When a gate is used and its duration is held constant, it turns out that less input power is required if the PRF is decreased. This is because the gating factor is decreased, which in turn means that the background noise is lowered with respect to the input signal level, and thus we are able to obtain greater sensitivity (see Eq. (79)). Theory shows (cf. Eq. (80)), that with a quadratic rectifier and a gate of constant duration, the mean power necessary to achieve a given signal-to-noise ratio is proportional to the square root of the PRF. (The peak power is correspondingly inversely proportional to the square root of the PRF.) This result is general and does not require a particular pulse or filter shape, such as, for instance, the gaussian. Curves giving the input power as a function of PRF are shown in Fig. 7.* Here it is assumed that the filter is matched and gaussian, but in other cases, the curves will not be appreciably different, in fact, not at all in the case of the quadratic rectifier.

Instead of keeping the gate fixed and varying the PRF, one may also be interested in what happens when the PRF is constant and the width τ_g of the gate is altered. The variation of the necessary input power with τ_g under these conditions is shown in Fig. 8. Obviously, narrowing the gate has the same effect as narrowing the earphone or meter filter. The abscissa of Fig. 8 can thus be equally well taken as the gate or the earphone width, depending on which one is interested in varying. For the quadratic detector, where questions of modulation suppression do not enter, the power is proportional to the square root of the gate width (or of the audio-filter

* The curves in Fig. 7 for the linear detector do not extend beyond a certain value of $\bar{P}$ for any particular σ_0 ($=s_F/\bar{P}$ at match). The reason for this is mathematical and not physical, for the formulas involved are not reasonably correct beyond the point indicated.

width). This behavior for the quadratic rectifier, which also applies to the linear detector with very weak signals, does not require any specialization, but the curves of Fig. 8 for the response of linear rectifiers with non-vanishing values of σ_{00} apply accurately only to a matched, gaussian pulse. The plots for other filters, however, will not be appreciably different.

When we seek to build up the efficiency by narrowing the gate, and when the latter precedes the IF filter, there are limits beyond which we cannot go. For if we were to decrease the gating factor to too narrow a relative time interval, it would be impossible to detect a target which was moving rapidly enough to pass through the equivalent space range of the gate between two successive scans, since the echoes returning from the target might then arrive at the receiver during a period in which the gate were shut. Apart from these considerations, gating under such circumstances would result in a distortion of the pulse by shearing it fore and aft (an effect not considered in our calculations), so that the condition for best match for the filter would be changed. If a gate of short duration is used at all, it is better to place it after the IF filter and corresponding rectification. As noted by North, it is possible

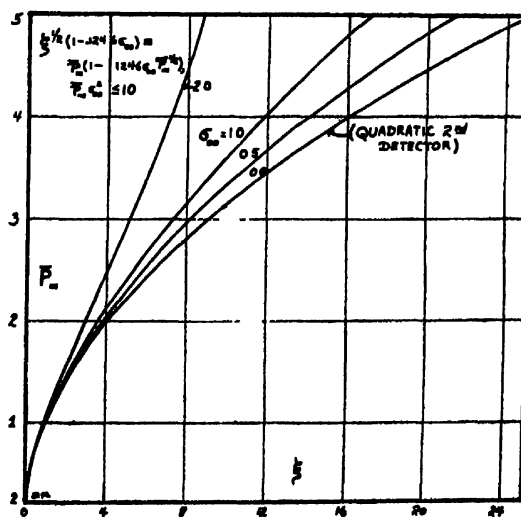


FIG. 8. Input power to the IF for aural or aperiodic meter reception of the minimum detectable signal at match using a linear rectifier, as a function of ξ , where ξ represents (1): duration of the gate, or (2) width of the audio-filter, or (3) frequency width of pulse (or IF filter), where gaussian pulses and filters are assumed. The subscript 00 refers to matching the IF to the pulse.

under these circumstances to use a gate of infinitesimal length. The item of adequate space coverage of targets is then secured by the finite length of the pulse. Of course, the ideal system would be to have an infinitesimal gate come right at the maximum of the rectified pulse, but uncertainty in range will in general make this impossible. The great advantage of visual detection is that the eye is, so to speak, its own properly timed gate. In attempting to discern a signal, it compares a pip or signal maximum only with the noise background in its immediate vicinity in time. North's rather elegant concept of an infinitesimal gate has the advantage that if the gate is inserted at just the right time, i.e., at the maximum of the pulse, the criterion for a minimum detectable signal is obviously similar for both aural and visual detection, for then in either case we are concerned only with the ratio of signal-to-noise energy at the very peak of the signal. In practise, however, an ideally-timed, infinitesimal gate seems too much of a simplification, for reasons already stated, but still it is a useful concept,* delineating a limiting behavior.

VII. Relative and Absolute Sensitivities of Visual and Aural Detection: Behavior at Match

As already mentioned many times, we say that the filter is "matched" if it is the conjugate of the Fourier transform of the pulse. The study of the matched case is particularly fundamental for three reasons (a), it is always a measure of best achievable performance, (b), it furnishes a convenient norm with which to compare results in other cases, and (c) it is particularly simple from the mathematical standpoint, especially in the visual case.

Hence in our attempts to calculate absolute sensitivities numerically in the end of this present

* In our mathematical section (Part II) we do not treat the infinitesimal gate. However, we can apply in a rather straightforward manner a fluctuation analysis, applied to pulses which are the delta-functions passed by this infinitesimal gate. If we assume that the gate is properly timed, and that a second gate is present to cancel out the discrete noise in the fundamental of the PRF, then it can be shown that if $(s/n)_P$ is given, $(s/n)_\infty$ in the gate exceeds the visual ratio $(s/n)_v$ by a factor $(T\Delta f_c)^{-1}$, where $(T\Delta f_c)^{-1}$ is the ratio of PRF to audio- (or meter) filter width. Even with perfect timing the performance is not appreciably improved by making the gate much shorter than the pulse. In other words, our later Eq. (c) of Section VII, with $\tau \sim 1/\Delta f_c$, gives a value of $(s/n)_v$ of the same order as that which we have just estimated for an infinitesimal gate.

section, we shall always assume that a condition of match exists. When in the present section we talk about the dependence on pulse length at match, the reader should be careful not to confuse this type of dependence on pulse length with that treated in the preceding section, where the filter width was assumed constant. We now consider the type of variation for which, when the pulse length is altered, the filter is correspondingly changed so as to be always at match.

With *visual perception* (on the deflection-modulated oscilloscope), the theory of Section III, Part II shows that the results obtainable at match are completely independent of the pulse shape (on the assumption of the "simple" criterion, of course). For the square detector one has, at match, from (67) and (73), the very simple result

$$(s_n/n_v)_{\text{quad}} = P_0 T_0 / W_0,$$

where W_0 is the power density of input noise, per unit frequency range, in the spectral region of the IF filter, P_0 is the input power, and T_0 is the pulse repetition period. For the linear detector the corresponding formula is, from (67) and (75)

$$(s_n/n_v)_{\text{lin}} = 0.957 \frac{P_0 T_0}{W_0} \times \left(1 - 0.1526 \left(\frac{P_0 T_0}{W_0} \right)^{1/2} \right), \quad P_0 T_0 / W_0 \leq 10.$$

Since P_0 and T_0 occur only in the product $P_0 T_0 / W_0$, the only thing that counts is the input energy per pulse, viz., $E = P_0 T_0$, Eq. (63). This is because our "simple" criterion for visual performance neglects the integrating effect of the eye, as already discussed in detail in Section VI, Part I.

Our conclusion that in visual detection the performance at match is independent of the pulse length is doubtless too much of an idealization, as the "simple" criterion takes no cognizance that the visibility of a pip is probably somewhat conditioned by its width. However, widening or narrowing of the pulse and pip will also be accompanied at match by fattening or shrinking of the "blades of grass" due to noise. Because of this fact, the invariance of pulse structure pre-

TABLE I.

	OUTPUT RATIOS		INPUT RATIOS σ_0			
	Physically minimum Detectable signal		Quadratic 2nd detector $\Delta f_s = \Delta f_n$ 50 c.p.s.	Quadratic 2nd detector Δf_s 5 c.p.s.	Linear 2nd detector Δf_s 50 c.p.s.	Linear 2nd detector Δf_s 5 c.p.s.
(Visual) (s_s/n_s)	0.0 db (pessimistic)		$\sigma_0 =$	0.0 db		+1.0 db
	- 5.0			-2.5		-1.7
	-10.0 (optimistic)			-5.0		-4.4
(aural) (s_s/n_s) no gate	+10.0 (pess.)		+13.5	+ 8.5	>13.5	+11.2
	0.0 (opt.)		+ 8.5	+ 3.5	>13.5	+ 4.8
(aural) (s_s/n_{s1}) gate	+10.0 (pess.)		+ 2.0	- 3.0	+ 2.9	- 2.5
	0.0 (opt.)		- 3.0	- 8.0	- 2.5	- 7.8
	Δf_n	Time constant $=2/\Delta f_n$				
periodic metering (s_m/n_m)=0.0 db	2 c.p.s.	1 sec.	$\sigma_0 = -6.7$ db	- 9.3	- 6.3	- 9.0
gated noise, lin. rect.	1/5	10	- 9.3	-12.0	- 9.0	-11.5
	1/10	20	-10.0	-12.4	- 9.8	-12.2
aperiodic metering (s_m/n_m)=0.0 db	2	1	-10.0			- 9.7
	1/5	10	-15.5			-15.2
	1/10	20	-16.5			-16.3

The equivalent rectangular pulse is 1.41 μ sec. long. $T_0 = 10^{-3}$ sec., $\tau_0 = 5 \times 10^{-3}$ sec., $f_0 = 4 \times 10^3$ c.p.s.

dicted by our "simple" criterion is probably not too much of a sin, unlike the conclusions which it yielded concerning the dependence of visibility on PRF.

With *aural reception* the performance at match is no longer independent of pulse shape or length. Instead, the power required to reach a minimum detectable signal-to-noise ratio is less, the longer the pulse. The dependence of this power on pulse duration turns out to be similar to that on gate length studied in the previous section. As long as the shape of the pulse remains constant, and only its time-scale is changed, the power demanded is inversely proportional to the square root of the length of the pulse, if the detector is quadratic or if it is linear and the signal is weak. The variation of the necessary power, with pulse length in arbitrary units, is shown in Fig. 8 for the quadratic detector and for the linear rectifier with various thresholds.

The reason for this difference in behavior between the visual and aural cases as regards the dependence on pulse length is that the expression to be maximized under optimum filter design is different in the two cases, and it can be shown that in consequence the maximum in one case has the value unity quite irrespective of the shape of the pulse, whereas in the other the maximum

depends on the assumptions concerning the structure and length of the pulse (cf. Eqs. (52), (55), and remarks in second paragraph after Eq. (56)). Physically the reason that increasing the pulse length is helpful in aural but not visual detection is that it lowers the fraction of energy wasted in irrelevant higher harmonics of the PRF in aural reception, whereas in the visual case it prolongs the maximum without raising it, which with our "simple" criterion is no help.

Numerical Calculations of Sensitivities

Knowing the values of the physically minimum detectable signal-to-noise ratios for the various types of presentation, oscilloscope, ungated, or gated aural reception, and use of a meter, we can calculate and compare the relative and absolute sensitivities of these methods. To do so we use the results obtained in Part II, and quote some of the final formulas here. For visual reception we have the relations

$$(s_s/n_s)_{\text{quad}} = \sigma_0^2; \quad (a)$$

$$(s_s/n_s)_{\text{lin}} = 0.957\sigma_0^2(1 - 0.1526\sigma_0), \quad \sigma_0 \leq 10^4,$$

which follow from (73) and (75), at match. For aural reception, the necessary formulas are, from (48a, b) and (49a, b) and Section IV (b), Part II

for ungated noise,

$$(s_a/n_a)_{\text{quad}} = \frac{0.633\sigma_0^2}{T_0(\Delta f_s f_b)^{\frac{1}{2}}};$$

$$\left(\frac{s_a}{n_a}\right)_{\text{lin}} = \frac{0.616\sigma_0^2}{T_0(\Delta f_s f_b)^{\frac{1}{2}}}(1 - 0.1246\sigma_0), \quad (b)$$

$$\sigma_0 \leq 10^{\frac{1}{2}}.$$

Here Δf_s is the width of the critical ear or audio-filter, whichever is narrower. When the background is gated, (b) becomes

$$(s_a/n_{aG})_{\text{quad}} = \frac{0.633\sigma_0^2}{(T_0\tau_G\Delta f_s f_b)^{\frac{1}{2}}};$$

$$(s_a/n_{aG})_{\text{lin}} = \frac{0.616\sigma_0^2}{(T_0\tau_G\Delta f_s f_b)^{\frac{1}{2}}}(1 - 0.1246\sigma_0), \quad (c)$$

$$\sigma_0 \leq 10^{\frac{1}{2}}.$$

The expression for the periodic meter is obtained from (97) and is

$$(s_m/n_m) = 1.156 \frac{[1 + \frac{1}{2}(\Delta f_m/\Delta f_s)^2]^{\frac{1}{2}}}{(\Delta f_m/\Delta f_s)^{\frac{1}{2}}}$$

$$\times \{(s_a/n_{a(G)})^2 - 0.1526(s_a/n_{a(G)})^2\},$$

$$s_a/n_{a(G)} \leq 10^{\frac{1}{2}}, \quad (d)$$

where the second detector that is used may be either linear or quadratic. The parenthesis around the subscript G in (d) means that the appropriate value of s_a/n_a is (b) or (c) according as the noise is ungated or gated.

The quantity Δf_m appearing in (d) is twice the equivalent rectangular filter width of the meter. (The reason for the factor two is that our general theory contemplates a filter extending symmetrically on the two sides of a resonance center, whereas the low pass filter involved in a meter is unilateral.)

To illustrate and compare the absolute sensitivities for the various methods Table I has been prepared for the following conditions of operation. First, we have assumed a gaussian pulse, to which the filter is matched. The width of the equivalent rectangular pulse is $7.1 \cdot 10^5$ c.p.s., $f_b = 4 \cdot 10^5$ c.p.s. from (60), and the width between half-power points is $\frac{1}{2}$ Mc. Furthermore, we have taken $\tau_G = 5 \cdot 10^{-6}$ second (the total duration of a double gate; see Sec. IV and VI), and $T_0 = 10^{-3}$

sec., corresponding to a PRF of 1000 c.p.s. Two different audio-filters are used: $\Delta f_s = 50$, $\Delta f_s = 5$ cycles. Three distinct meter filters are also employed: $\Delta f_m = 2$, $\frac{1}{2}$, $\frac{1}{10}$ cycles (actual width 1 cycle, $\frac{1}{10}$, $1/20$ cycles), each with the same audio-filter $\Delta f_s = 50$ or $\Delta f_s = 5$ c.p.s.

Here σ_0 equals (s_F/n_F) at match, and is given in db in the appropriate boxes of the table. We have assumed that $s_m/n_m = 1$, but this is only a guess at present, for want of more complete experimental data. From this table we see that the linear detector is always a bit worse than the quadratic, though the difference for weak signals is small. We notice also that under the above conditions of operation, aural reception is about as good as visual, and in some cases surpasses it, although the psychological strain of aural work may add as much as 5 db to these results. The meter, of course, compares favorably on the assumption that $s_m/n_m = 1$ as a minimum detectable signal. Narrowing the audio-filter by a factor 10 increases the sensitivity by 5 db. Eaton and Wolff (RCA Princeton, in an unpublished paper (Report PTR-7C) completed June 4, 1943), using the aperiodic method, with time-constants of the order of 10 seconds, obtained in the Laboratory a 20-db improvement over visual detection by the (deflection-modulated) oscilloscope if $s_v/n_v = 1$. This is about the performance that our theory indicates. Namely, if we take $\Delta f_m = \frac{1}{2}$ c.p.s., corresponding to a time-constant of 10 seconds, and take a single gate of about 1.5μ -seconds duration, as they used a gate comparable with the pulse length (rather than the longer and double gates assumed in our tables), our formulas yield an 18.5 db gain in sensitivity over the "pessimistic" oscilloscope reading. We cannot, however, emphasize too much that the efficacy of the various metering methods is contingent upon the feasibility of using a long time-constant, and this will depend on tactical considerations (of Sec. IV).

PART II. MATHEMATICAL SECTION

I. Introduction: Signal and Noise in the IF Filter

The pulse-modulated carrier enters the RF stages and the carrier frequency is shifted to the lower frequency, (IF), by means of the first detector or mixer, as it is often called. We assume

that this displacement of the carrier, as well as the performance of the RF stages, is essentially a linear process, resulting in no appreciable distortion of the incoming wave. Our real problem, then, is with the IF and subsequent stages of the receiver. The noise entering the IF arises primarily in the mixer, or in the case of artificial noise, comes, of course, through the RF. The noise and signal are filtered before entering the second detector. After rectification, the mixture is filtered by the video- or an audio-filter, amplified, and passed finally to an oscilloscope, earphones, or to a meter.

The signal entering the IF stages has the following forms:

$$f_I(t) = \sum, \sqrt{2} s_I(t-t_j) \cos \omega_0(t-t_j), \quad (1)$$

where $\omega_0 = 2\pi f_0$ and f_0 is the intermediate frequency. Equation (1) represents the incoming wave as a succession of pulses, the envelope of any one of which is described by $\sqrt{2} s_I(t-t_j)$. The different values of t_j differ by the pulse repetition period T . A single disturbance is expressed by $\sqrt{2} s_I(t) \cos \omega_0 t$, with $t_j = 0$ for convenience. We shall also assume that the overlapping between consecutive pulses is negligible, so that in determining (1) it is necessary to consider only one value of t_j at a time. For all purposes this is a very good approximation. The Fourier transform of a single pulse is then

$$S_I'(\omega') = \int_{-\infty}^{\infty} \sqrt{2} s_I(t) \exp(-i\omega' t) \cos \omega_0 t dt, \quad \omega' = 2\pi f'. \quad (2)$$

Note that from (2) it follows automatically that $S_I'(\omega') = S_I'(-\omega')^*$ and $S_I'(\omega')^* = S_I'(-\omega')$, where the star (*) denotes the complex conjugate. The corresponding relation to (2) is

$$\sqrt{2} s_I(t) \cos \omega_0 t = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_I'(\omega') \exp(i\omega' t) d\omega'. \quad (3)$$

As well as amplifying, the IF stages act like a filter, whose *amplitude* frequency response we denote by $G_{IF}'(\omega')$. For the time being we will not restrict the generality of our treatment by specifying $G_{IF}'(\omega')$ explicitly. The filtered signal that leaves the IF is then

$$s_F'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_I'(\omega') G_{IF}'(\omega') \exp(i\omega' t) d\omega', \quad (4)$$

where we must have $G_{IF}'(-\omega') = G_{IF}'(\omega')^*$ in order that we have a real output for a real input. The primes are dropped from the notation for the various filter characteristics when the origin for frequency is taken at the carrier f_0 rather than at the absolute zero of frequency: thus $\omega = \omega' - \omega_0$, $S_I'(\omega' - \omega_0) = S_I(\omega)$, $G_{IF}'(\omega' - \omega_0) = G_{IF}(\omega)$. Then with the substitution $\omega = \omega' - \omega_0$ we may write (2) without appreciable error as

$$S_I(\omega) = 2^{-1} \int_{-\infty}^{\infty} s_I(t) e^{-i\omega t} dt, \quad (5)$$

since we can disregard the part of the integrand with the oscillatory factor $\exp(-2i\omega_0 t)$, whose effect is negligible because of the rapid variation. Equation (4) can be written

$$s_F'(t) = 2 \times \text{Real Part of}$$

$$\left\{ \frac{\exp(i\omega_0 t)}{2\pi} \int_{-\infty}^{\infty} S_I(\omega) G_{IF}(\omega) e^{i\omega t} d\omega \right\}. \quad (6)$$

We can express the filtered wave as

$$s_F'(t) = \sqrt{2} s_F(t) \cos(\omega_0 t - \epsilon), \quad (7a)$$

where

$$\sqrt{2} s_F(t) = 2 \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} S_I(\omega) G_{IF}(\omega) e^{i\omega t} d\omega \right|. \quad (7b)$$

The phase factor ϵ is really a function of t but is so slowly varying that to all intents and purposes it can be regarded as constant. It is to be noted that the integrand of (7b) has a single resonance center at $\omega = 0$, whereas that of (4) has resonances at $\omega = \pm\omega_0$. In extending the lower limit from $-\omega_0$ to $-\infty$ in (7b) it is to be understood that $G_{IF}(\omega)$ is to be taken as very small for $\omega < -\omega_0$, whereas the literal definition $G_{IF}(\omega) = G_{IF}'(\omega' - \omega_0)$ would imply that $G_{IF}(\omega)$ had a resonance at $-2\omega_0$ also.

Equation (7b) shows that the amplitude $s_F(t)$ after filtering can be computed, as we would expect, just as though we ignored the carrier entirely, and applied the Fourier analysis directly to the modulation factor $s_I(t)$ of the incoming wave. We call $s_I(t)$ a modulation factor rather than an envelope because in evaluating (7b) it is essential to take account the algebraic sign as well as the modulus of $s_I(t)$.

We will assume that the background noise entering the IF is sensibly uniform over the filter width, having a constant spectral density W_0 (power/frequency). This is really no assumption at all, as the spectral variations in the noise over the IF filter width are usually of little account. The mean square noise power transmitted by the IF stage is then

$$n_F^2 = \frac{1}{2} \frac{W_0}{2\pi} \int_{-\infty}^{\infty} |G_{IF}'(\omega')|^2 d\omega' \\ = \frac{W_0}{2\pi} \int_{-\infty}^{\infty} |G_{IF}(\omega)|^2 d\omega, \quad (8)$$

where n_F is the r.m.s. noise voltage entering the rectifier.

II. Quadratic and Linear Rectification of Signal and Noise

We introduce now the correlation function $R(t)$ for the noise and signal output of a linear or non-linear device, such as a filter or rectifier, in terms of the power spectrum of the output $W(\omega)$. The correlation $R(t)$ is defined as the mean value of the product of the noise and signal amplitudes at times separated by an interval t , averaged over all initial phases. It can be shown⁸⁻¹⁰ also that the correlation function is the Fourier transform of the output power spectrum, and *vice versa*. We have thus the following general relations:

$$W(\omega) = 4 \int_0^{\infty} R(t) \cos \omega t dt, \quad (9)$$

and

$$R(t) = \frac{1}{2\pi} \int_0^{\infty} W(\omega) \cos \omega t d\omega. \quad (10)$$

For noise leaving the IF filter we have the

⁸ N. Wiener, Acta. Math. 55, 117 (1930), and later, independently, A. Khintchine, Math. Ann. 19, 604 (1934).

⁹ G. I. Taylor, Proc. Lond. Math. Soc. Sec. 2, 20, 196 (1920), and Proc. Roy. Soc. 164, 476 (1938).

¹⁰ S. O. Rice, Bell Sys. Tech. J. 23, 282 (1944), and 24, 46 (1945). These two papers contain an excellent and quite comprehensive treatment of a large number of noise problems.

¹¹ J. H. Van Vleck, "The Spectrum of Clipped Noise" (Harvard Radio Research Laboratory, Report 411-51, July 21, 1943). Available at Harvard War Archives, Littauer Building, Harvard University.

¹² D. Middleton, "Note on The Theory of Square Law Rectification of Modulated Carrier in Presence of Noise." (Harvard Radio Research Laboratory, completed July 12, 1944). Available at Harvard War Archives, Littauer Building, Harvard University.

correlation function $\psi(t)$, viz.

$$\psi(t) = \frac{W_0}{4\pi} \int_{-\infty}^{\infty} |G_{IF}'(\omega')|^2 \cos \omega' t d\omega'. \quad (11)$$

A comparison of (11) and (8) shows immediately that $n_F^2 = \psi(0)$, so that $\psi(0)$, or in abbreviated form ψ , is the mean square noise voltage passed by the IF filter, which is centered about f_0 . Then with the help of (11) we can write with negligible error

$$\psi(t) = \psi_0(t) \cos \omega_0 t \\ = \psi(0) r(t) = \psi(0) r_0(t) \cos \omega_0 t. \quad (12)$$

The relations (12) define the normalized correlations $r(t)$ and $r_0(t)$, where

$$\psi_0(t) = \frac{W_0}{2\pi} \int_{-\infty}^{\infty} |G_{IF}(\omega)|^2 \cos \omega t d\omega, \quad (13)$$

since the contribution to the integral when $\omega < -\omega_0$ is ignorable. Note that $r_0(0) = r(0) = 1$.

(a). Noise and Signal Output of a Quadratic Rectifier

Let us now pass the signal from the IF with its accompanying noise through a square-law detector. It can be shown¹⁰ that the correlation function $R_N(t)$ before averaging over the phases of the modulation $s_F(t)$ is

$$R_N[t, t_0, t_0'] = \eta^2 \{ \psi^2 + b_1^2 b_2^2 \psi^2 + 2\psi(t)^2 \\ + 4b_1 b_2 \psi(t) + b_1^2 \psi^2 + b_2^2 \psi^2 \}, \quad (14)$$

where

$$b_1 = \sqrt{2} s_F(t_0') \cos \omega_0 t_0' / \psi^{\frac{1}{2}}; \\ b_2 = \sqrt{2} s_F(t_0' + t) \cos \omega_0 (t_0' + t) / \psi^{\frac{1}{2}}, \quad (15)$$

and η is a constant depending on the dynamic characteristic of the particular rectifier. Assuming that the envelope $s_F(t_0')$, etc., is a slowly varying function of the time, we may treat it as essentially constant, and hence uncorrelated with $\cos \omega_0 t_0$, etc., when averaging over the phases of the carrier. Then analytically t_0 and t_0' are independent. Now in the final form of the correlation function the terms in ψ^2 , $b_1^2 \psi^2$, $b_2^2 \psi^2$ contribute only to the d.c. Further, the term in $\psi(t)^2$ with the help of (12) may be resolved into $\psi(t)^2 = \frac{1}{2} \psi_0(t)^2 (1 + \cos 2\omega_0 t)$, the latter por-

tion of which contributes to the spectral region centered about $2f_0$, well removed from the region of our interest. Accordingly, the low frequency correlation function derived from (14) becomes in final form¹¹

$$R_{LF}(t) = \eta^2 \{ \overline{\psi_0(t)^2} + 2\overline{\psi_0(t)s_F(t_0')s_F(t_0'+t)} + \overline{(s_F(t_0')^2s_F(t_0'+t)^2)} \}, \quad (16)$$

where the bars indicate the average over t_0' . The first term of (16) represents noise alone, the second is the contribution to the continuum from the cross-modulation of the signal and the noise, and the third is the signal output alone. In both visual and aural reception we can in general disregard the middle, or second, term. The reasons behind this, however, are quite different in the two cases. For aural performance the signal is on such a small fraction of the pulse period that the contribution of the cross-term to the continuum is negligible in comparison with the noise and the signal terms. In the video case it is still true that these cross-terms are present only a small fraction of the time, but now it is that fraction of the time that interests us. However, the response is always a monotonically increasing function of s_F/n_F , which is physically obvious, since increasing the signal must always lead to an improvement in reception. The best performance is thus obtained by maximizing s_F/n_F . We may then take as our definition of signal to noise ratio the r.m.s. signal (envelope) at that point in the receiver circuit where we perceive it, to the r.m.s. value of the corresponding noise background, omitting the cross-modulation terms.

In the absence of signal the low frequency correlation function is, from (16),

$$R_{LF}(t) = \eta^2 \psi(0)^2 r_0(t)^2 = \eta^2 \psi_0(t)^2, \quad (17)$$

and from (10) it follows when $t=0$, that the low

¹¹ Equations (14) and (16) are derived on the assumption of small signal quadratic detection, arising from the square-law character of the dynamic response in the neighborhood of the operating point. It often happens that "large-signal" rectification occurs, where now the rectifier may have essentially a half-wave quadratic response. However, the results for the low frequency correlation are proportional, Eq. (16) being 4 times as great as the corresponding expression for half-wave detection. This is easy to see when we observe that only the envelope of the incoming, narrow-band disturbance is reproduced, albeit squared, in the low frequency output, exclusive of d.c. The factor of proportionality follows since in the latter case but half the input wave is transmitted.

frequency mean power output is

$$R_{LF}(0) = \eta^2 \psi(0)^2 = \eta^2 n_s^2 = \eta^2 n_F^4, \quad (18)$$

where n_s is the mean square noise voltage in the video, provided the effective band width of the IF is less than that of the video stages, so that the low frequency output after detection is not appreciably affected by the video response. The spectrum of the low frequency noise from the quadratic rectifier may be obtained from (9) with the help of (17), and is*

$$W_{LF}(\omega) = 4 \int_0^\infty R_{LF}(t) \cos \omega t dt. \quad (19)$$

We have seen that (18) gives an expression for the video noise background. In aural reception we are interested in detecting a single component against a relatively narrow section of the spectrum, so that it is the spectral level of the noise near $f=0$ that concerns us. From (19) this is

$$W_{LF}(0) = 4 \int_0^\infty R_{LF}(t) dt. \quad (20)$$

The mean square aural background noise n_a^2 is rigorously

$$n_a^2 = \frac{1}{2\pi} \int_0^\infty W_{LF}(\omega) |G_a(\omega)|^2 d\omega, \quad (21a)$$

where $G_a(\omega)$ is the amplitude response of the audio-filter, but we can use the approximation

$$n_a^2 = \frac{W_{LF}(0)}{2\pi} \int_0^\infty |G_a(\omega)|^2 d\omega. \quad (21b)$$

It is convenient to express the band width of physical filters, whether measured from the half-power or other points of the response, in terms of a rectangular filter whose spectral ordinate is equal to the maximum response of the particular filter in question, and whose width is such that both devices transmit equal noise power, i.e., such that the areas under the respective responses are the same. (We assume filter characteristics with but a single maximum; the criterion may be generalized for more involved spectral shapes.)

* Note: After rectification there exists no carrier frequency in the spectral region with which we are concerned; here ω refers to absolute frequency, and not to frequency measured with respect to ω_0 .

If $G_k(f)$ is the response of the physical filter, then the effective band width Δf_k of the equivalent rectangular apparatus becomes

$$\Delta f_k = \int_{-\infty}^{\infty} |G_k(f)|^2 df / |G_k(f)|_{\max}^2. \quad (22)$$

In all our definitions we will assume that $|G_k(f)|$ is normalized so that $|G_k(f)|_{\max} = 1$. Then (21b) may be written still more simply as

$$n_s^2 = W_{LF}(0) \Delta \omega_s / 2\pi = W_{LF}(0) \Delta f_s, \quad (23)$$

where in this instance Δf_s is the width of the earphone or audio-filter, if the band is very narrow, or the actual band width of the ear, if the audio-filter is wider. (See Fig. 1 for curves of (Δf_s) critical.)

(b). Noise Output of a Linear Rectifier

For the linear detector the low frequency part of the correlation function exclusive of d.c. for noise alone after rectification may be shown to be¹²

$$\begin{aligned} R_{LF}(t) &= \frac{\beta^2 \psi(0)}{2\pi} \{ {}_2F_1(-\tfrac{1}{2}, -\tfrac{1}{2}; 1; r_0(t)^2) - 1 \}, \\ &= \frac{\beta^2 \psi(0)}{8\pi} \{ r_0^2 + r_0^4/16 + r_0^6/64 \\ &\quad + 25r_0^8/4096 + \dots \} \\ &= \frac{\beta^2 \psi(0)}{8\pi} \sum_{n=0}^{\infty} \left\{ \frac{(2n)!}{2^{2n} n! (n+1)!} \right\}^2 r_0^{2n+2}, \\ &= \frac{\beta^2 \psi(0)}{2\pi^2} \{ 4E(r_0) - 2(1-r_0^2)K(r_0) - \pi \}, \end{aligned} \quad (24)$$

where ${}_2F_1$ is the (Gaussian) hypergeometric function, and E and K are complete elliptic

¹² The various forms of $R_{LF}(t)$ appear to have been derived independently by Franz, *Zeits. f. Hochfrequenz-technik*, p. 140 (1941), who gave essentially the series expression, by D. O. North, who mentions it in a synopsis of a paper entitled "The Modification of Noise by Certain Non-Linear Devices," presented at the Winter Technical Meeting of the I.R.E., Jan. 28, 1944, and by G. E. Uhlenbeck in terms of the complete elliptic integrals E and K , in (M.I.T.) Radiation Laboratory Report 453, "Theory of Random Processes," Oct. 15, 1943. These results may also be obtained from Rice's expression (4.7-6), reference 8, for the correlation function of the noise output after half-wave linear rectification, with the help of (12) and the expansion of $(\cos \omega_0 t)^{2n}$.

integrals of modulus r_0 , with $r_0(t)$ specified by (12). The mean square video noise voltage n_s^2 without signal follows at once from (24), with the aid of

$${}_2F_1(\alpha, \beta; \gamma; 1) = \Gamma(\gamma) \Gamma(\gamma - \alpha - \beta) / \Gamma(\gamma - \alpha) \Gamma(\gamma - \beta),$$

$$R(\gamma) > 0, \quad R(\gamma - \alpha - \beta) > 0.$$

We have

$$n_s^2 = \beta^2 \psi(0) \left[\frac{2}{\pi^2} - \frac{1}{2\pi} \right] = 0.0435 \beta^2 \psi(0). \quad (25)$$

The low frequency continuum to be used in (21), or in (23), is immediately obtained from (19) and (24), to wit:

$$W_{LF}(\omega) = \frac{\beta^2 \psi}{2\pi} \sum_{n=0}^{\infty} \left\{ \frac{(2n)!}{2^{2n} n! (n+1)!} \right\} \times \int_0^{\infty} r_0(t)^{2n+2} \cos \omega t dt. \quad (26)$$

(c). Signal Output of a Quadratic Rectifier in Visual and Aural Reception

As can be seen from (16), the low frequency signal output of a quadratic detector is independent of the background noise, if we neglect the cross-modulation term; (see reference 11). In visual reception it is the mean value of the area under one cycle of the IF carrier, or equivalently, the envelope of the pulse, that is important. The signal displayed appears as a single pulse $s_s(t_0' - t_s)$. We have assumed that the modulation is very slow compared to f_0 , and so $s_s(t_0' - t_s)$ is essentially constant in time, t , when it appears in $R_N(t, t_0, t_0')$, as in integrating over t_0 we are dealing with times of the order of $1/f_0$. In order to emphasize the approximate constancy of $s_s(t_0' - t_s)$ in this integration we denote the argument by $t_0' - t$, rather than by $t_0 - t_s$. However, it is not legitimate to ignore the time dependent nature of $s_F(t)$ and $s_s(t)$ when later we consider signal variations over times of the order of ω_p^{-1} , where ω_p is the angular frequency, or PRF. From (16) we find that the mean pulse power is

$$R_{LF}(0) = \langle s_s(t_0')^2 \rangle_m = \eta^2 \langle s_F(t_0')^4 \rangle_m, \quad (27)$$

and for the signal output we have

$$s_s(t_0') = \eta s_F^2(t_0'). \quad (28)$$

In half-wave quadratic rectification we need merely divide (27) and (28) by 4. For video responses, then, (28) is simply

$$s_s = \eta s_F^2, \quad (29)$$

where s_s and s_F are the maximum values of $s_s(t_0')$ and $s_F(t_0')$, respectively. Hitherto we have denoted the time argument of the signal amplitude by t_0' , distinguishing it from the argument t_0 of the carrier oscillations so that in integrating over the latter we could treat t_0' as constant. Since this integration has been performed, the prime can now be dropped.

To obtain the audio-spectrum we must take the Fourier components of the modulation. Now in aural reception we listen to a single component of the pulse, corresponding to the PRF or a higher harmonic thereof. Replacing t_0' by t in (28) and expanding $s_s(t)$ in a Fourier series, we find

$$s_s(t) = \sum_{-\infty}^{\infty} a_k \exp(ik\omega_p t), \quad \omega_p = 2\pi f_p = 2\pi/T, \quad (30)$$

where T is the period of the pulses and the amplitude of the k th harmonic of the PRF f_p is given by

$$a_k = \left\{ \frac{1}{T} \int_{-\infty}^{\infty} s_s(t) e^{-i\omega_k t} dt \right\}_{\omega = k\omega_p}, \quad (31)$$

since successive pulses are assumed not to overlap appreciably. Equation (30) may also be written as

$$s_s(t) = B_0 + \sum_{k=1}^{\infty} B_k \cos(k\omega_p t - \epsilon_{kp}), \quad (32)$$

where ϵ_{kp} is an uninteresting phase factor and

$$B_0 = \left| \frac{1}{T} \int_{-\infty}^{\infty} s_s(t) e^{-i\omega t} dt \right|_{\omega=0}; \quad (33)$$

$$B_k = \left| \frac{2}{T} \int_{-\infty}^{\infty} s_s(t) e^{-i\omega_k t} dt \right|_{\omega = k\omega_p}.$$

If $\omega_k = \omega_p k$ remains small compared with the IF band width, we may set $\omega = 0$ in (33) because then the trigonometric factor varies but little during the time the pulse is on, and we may take $\sqrt{2}B_0$ as the r.m.s. amplitude s_s of the k th component we are listening to. With the help of (28), Eq. (33) becomes then

$$B_k = s_s = \eta \frac{\sqrt{2}}{T} \int_{-\infty}^{\infty} s_F(t)^2 dt. \quad (34)$$

(d). *Signal Output of a Linear Detector in the Presence of Noise: Visual and Aural Reception*

The expression for the shape of the signal wave transmitted by a linear rectifier is complicated, because it is quite different from what it would be in the absence of noise, the interference from the latter giving rise to the phenomenon which is known as "modulation suppression." This phenomenon results in a suppression of the signal, or of the noise, whichever is small compared with the other. The detailed mathematical theory shows that the low frequency part of the rectified signal is given by^{1,13-15}

$$s_s(t_0') = \frac{\beta n_F}{(2\pi)^{1/2}} {}_1F_1\left(-\frac{1}{2}; 1; -s_F(t_0')^2/n_F^2\right), \quad (35)$$

where ${}_1F_1$ is a confluent hypergeometric function and β , as before, is the dynamic transconductance of the rectifier. The function ${}_1F_1$ is specified in general by the series

$${}_1F_1(\alpha; \beta; x) = 1 + \frac{\alpha x}{\beta 1!} + \frac{\alpha(\alpha+1)x^2}{\beta(\beta+1)2!} + \dots + \frac{\alpha_n x^n}{\beta_n n!} + \dots \quad (36)$$

The video response $s_s(t_0')$ represents the mean area under a cycle of the IF carrier, times the envelope of the modulation. In comparing signal and noise we must be careful to define what we mean by a signal and by noise, and to measure them in the same way. Hence, since noise is measured as an r.m.s. voltage, we must determine the signal similarly. This is the reason for taking the mean area under a cycle of the IF carrier as the signal criterion, rather than the envelope, a point which has not always been brought out in previous treatments of the subject.

Equation (35) as it stands includes both signal and d.c. noise; to obtain the increment in video

¹³ See S. O. Rice, reference 8, Eq. (4.2-3).

¹⁴ S. A. Goudsmit, "Comparison Between Signal and Noise" (M.I.T. Radiation Laboratory, Report No. 193, January 29, 1943).

¹⁵ W. H. Jordan, "Action of Linear Detector on Signals in the Presence of Noise" (M.I.T. Radiation Laboratory, Report No. 305, July 6, 1943).

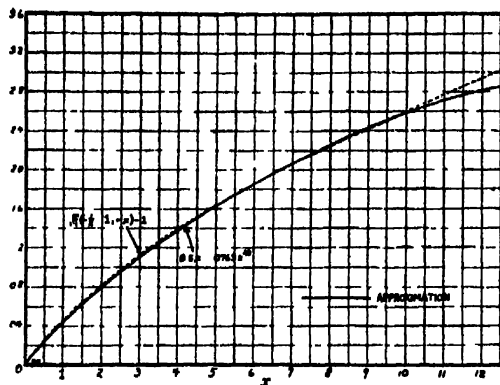


FIG. 9. Approximation of ${}_1F_1(-\frac{1}{2}; 1; -x) - 1$ by $\alpha x - \gamma x^k$, $\alpha = \frac{1}{2}$, $\gamma = 0.0763$, $k = 1.50$.

response due to the signal we must subtract the mean d.c. response in the absence of the signal, thus obtaining

$$s_v(t'_0) = \frac{\beta n_F}{(2\pi)^{\frac{1}{2}}} \times \{ {}_1F_1(-\frac{1}{2}; 1; -s_F(t'_0)^2/n_F^2) - 1 \}, \quad (37)$$

inasmuch as ${}_1F_1(\alpha; \beta; 0) = 1$. At the instant of maximum signal, which is important in visual detection, we have by definition $s_v(t'_0) = s_v$, $s_F(t'_0) = s_F$, and Eq. (37) becomes

$$s_v = \frac{\beta n_F}{(2\pi)^{\frac{1}{2}}} \{ {}_1F_1(-\frac{1}{2}; 1; -s_F^2/n_F^2) - 1 \}, \quad (38)$$

where as before s_F is the maximum value of $s_F(t')$. When the signal is weak compared with the noise we find that (38) is

$$s_v \approx \frac{\beta}{2(2\pi)^{\frac{1}{2}}} n_F \left(\frac{s_F}{n_F} \right)^2, \quad (s_F/n_F)^2 \ll 1, \quad (39)$$

which is a condition often encountered in practice. On the other hand, when the signal is very large compared to the noise, or what is the same thing, for a finite signal and negligible noise, we have with the help of the asymptotic expansion of ${}_1F_1$

$$s_v \approx \frac{\beta\sqrt{2}}{\pi} s_F, \quad (s_F/n_F)^2 \gg 1, \quad (40)$$

which, as one would expect, is $1/\pi$ times the envelope of the IF wave.

In aural reception we listen to a single component of (37). With the aid of (30)–(34) we

may write for the r.m.s. value s_a of this component

$$s_a = \frac{\beta n_F}{T\pi^{\frac{1}{2}}} \int_{-\infty}^{\infty} \{ {}_1F_1(-\frac{1}{2}; 1; -s_F(t)^2/n_F^2) - 1 \} dt. \quad (41)$$

Although the power series for ${}_1F_1(-\frac{1}{2}; 1; -x)$ is absolutely convergent for all values of x , it does not converge rapidly enough for computation when x is large, say $x > 1.5$. This difficulty may be overcome by approximating in a purely *ad hoc* fashion the integrand of (41) with a function of the form

$$F(x) = \alpha x - \gamma x^k. \quad (42)$$

Here α is the derivative of ${}_1F_1$ at $x=0$, and γ and k are determined from ${}_1F_1$ and its first derivative at some point $x=x_0$ which represents the maximum value of $s_F(t)^2/n_F^2$. We have then

$$\begin{aligned} \alpha &= {}_1F_1'(0) = \frac{1}{2}, \\ \text{where} \quad F'(x_0) &= \frac{1}{2} {}_1F_1(\frac{1}{2}; 2; -x_0), \\ F(x_0) &= {}_1F_1(-\frac{1}{2}; 1; -x_0) - 1, \\ \gamma &= [\alpha x_0 - F(x_0)] x_0^k, \\ k &= \left(\frac{F'(x_0) - \alpha}{F(x_0) - \alpha x_0} \right) x_0. \end{aligned} \quad (43)$$

The necessary values of the hypergeometric functions may be obtained from Jahnke and Emde.¹⁶ The approximation (42) requires that for each value of x_0 we calculate anew γ and k ; however, in practice the error is not great if we select some value of x_0 , say 6, and use the γ and k obtained in this case for all x_0 in the range $0 \leq x_0 \leq 10$. Figure 9 compares our approximation with the correct curve. The difference in the region $0 \leq x_0 \leq 10$ is seen to be quite small. Thus for subsequent work we choose $x_0 = 6$ and with this value find from (43) that

$$\gamma = 0.0763; \quad k = 1.50; \quad x_0 = 6. \quad (44)$$

Then the r.m.s. audio-component (41) after linear rectification in the presence of noise becomes finally:

$$s_a = \frac{\beta n_F}{2\pi^{\frac{1}{2}} T} \int_{-\infty}^{\infty} \{ s_F(t)^2/n_F^2 - \gamma s_F(t)^{2k}/n_F^{2k} \} dt. \quad (45)$$

¹⁶ Jahnke and Emde (Stechert, New York, 1938), pp. 275 *et seq.*

(e) Comparison of Signal with Noise in Visual and Aural Reception

When we listen or look for a signal with the help of our receiving apparatus, we do so against a noise background, man-made or natural, which tends to confuse or obscure our perception. Our ability to detect a signal will then obviously depend on the magnitude of the ratio of the signal at the point of reception (i.e., on the screen, at the carphones, or on a meter), to the appropriate noise background. In Sections (a)–(d) the various signal and noise voltages that apply in our problem have been determined; it remains now to compare them.

For visual reception involving quadratic detection we have from (18) and (29)

$$[s_v/n_v]_{\text{quad}} = (s_F/n_F)^2. \quad (46)$$

For linear rectification this ratio is modified to

$$[s_v/n_v]_{\text{lin}} = \rho_0 [(s_F/n_F)^2 - 2\gamma(s_F/n_F)^{2k}], \quad (47)$$

$$\rho_0 = \frac{1}{2}(4/\pi - 1)^{-1} = 0.957,$$

with the aid of (25), (38), and (42). When the signal is small in comparison with the r.m.s. noise, i.e., when $(s_F/n_F)^2 \ll 1$, we may neglect the second term in (47), and then the linear rectifier behaves almost exactly like the quadratic. In fact, for most practical purposes we may set ρ_0 equal to unity, so that then the performance of both types of detector is identical. For strong signals, $s_F/n_F \geq 1$, the quadratic rectifier becomes noticeably superior, i.e., (> 2 db), to the linear one, as shown by curve (a) of Fig. 2, and as described more fully in Sec. III following.

In aural work we find from (20), (23), and (34)

$$(s_a/n_a)_{\text{quad}} = \frac{1}{T(2\xi_0\Delta f_c)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{s_F(t)^2}{n_F^2} dt, \quad (48a)$$

with

$$\xi_0 = \int_0^{\infty} r_0(t)^2 dt. \quad (48b)$$

The corresponding expression for the linear rectifier becomes, from (23), (26), and (45):

$$(s_a/n_a)_{\text{lin}} = \frac{\rho_0'}{T(2\Delta f_c)^{\frac{1}{2}}} \times \int_{-\infty}^{\infty} \{s_F(t)^2/n_F^2 - 2\gamma(s_F(t)/n_F)^{2k}\} dt, \quad (49a)$$

where now

$$\rho_0' = \left[\sum_{n=0}^{\infty} \left(\frac{(2n)!}{2^{2n}n!(n+1)!} \right)^2 \int_0^{\infty} r_0(t)^{2n+2} dt \right]^{-1}. \quad (49b)$$

Again if the signal-to-noise ratio is small we may omit the second term in the numerator of (49a). Furthermore, there is no serious error in discontinuing the infinite series in (49b) after the first term. Then we have $\rho_0' = \xi_0^{-1}$, with the result that (49a) and (48a) become identical. In a similar approximation we can replace the series involved in (24) and (25) by their first terms, and then the constant of proportionality ρ_0 in (47) becomes ξ_0^{-1} instead of $0.957\xi_0^{-1}$, so that there is a like identity of results for the two detectors in the visual case.

When the series in (49b) is not replaced by its first term, the value of ρ_0' depends slightly on the filter characteristics, being somewhere between $0.957\xi_0^{-1}$ and ξ_0^{-1} . In the Gaussian case the value of ρ_0' is $0.973\xi_0^{-1}$. For practical purposes the difference in behavior of the two detectors is important only when we must consider the second term of the integrand of (49a). This makes the linear detector again less effective than the quadratic for strong signals (cf. curve (b) of Fig. 2). For the specific case of a Gaussian pulse and filter, consult Sec. IV, Part II.

III. Proof of the Fourier Transform Criterion

(a). Visual Reception

When the signal is detected visually, the best filter is by definition, according to our "simple" standards (see Sec. I, Part I), that which gives the largest value of the ratio s_v/n_v of peak video signal to r.m.s. video noise. This is equivalent to demanding that the ratio s_F/n_F of the maximum signal amplitude to r.m.s. noise amplitude after filtering, but before rectification, be as large as possible. By the maximum signal amplitude we mean the greatest absolute value of the "modulation factor" $s_F(t)$ for the signal $\sqrt{2}s_F(t) \cos \omega_0 t$ which is passed by the filter. The quantities s_v, n_v are by no means the same as s_F, n_F , respectively. For instance, the mean video noise power n_v^2 is but a fraction of the total noise n_F^2 passing through the filter, with the fraction depending on the type of detector, but not on the filter

design (cf. especially Eqs. (18) and (25)). However, the preceding analysis has shown that s_v/n_v is a monotonically increasing function of s_F/n_F . Namely, Eq. (46) shows that for the quadratic rectifier s_v/n_v is proportional to s_F^2/n_F^2 , as might have been guessed. For the linear detector s_v/n_v is proportional to s_F^2/n_F^2 only if $s_F^2/n_F^2 \ll 1$, but the hypergeometric function ${}_1F_1(-\frac{1}{2}; 1; -s_F^2/n_F^2)$ possesses the monotonic property for the entire domain in which we are interested. Hence maximization of s_v/n_v also implies maximization of s_F/n_F .

It was shown in Section I, Part II, that $s_v(t)$ can be computed, as we might suspect, by forgetting the carrier entirely, and regarding the origin of filter response as at the carrier rather than at zero absolute frequency. Thus if the signal entering the (IF) filter is $\sqrt{2}s_I(t) \cos \omega_0 t$, the value of the amplitude function $s_F(t)$ after filtering is, from (7b), at any given time t_0

$$s_F(t_0) = \frac{1}{\sqrt{2}} \left| \frac{1}{\pi} \int_{-\infty}^{\infty} G_{IF}(\omega) S_I(\omega) \exp(i\omega t_0) d\omega \right|. \quad (50)$$

The noise energy passed by the filter is given by (11) when t is set equal to zero. The input energy per pulse is

$$E = \int_{-\infty}^{\infty} s_I(t)^2 dt = \frac{1}{\pi} \int_{-\infty}^{\infty} |S_I(\omega)|^2 d\omega, \quad (51)$$

the second relation following in virtue of Plancherel's theorem¹⁷ and Eq. (5). Since (51) is independent of filter design the ratio which we desire to maximize can be regarded equally well as $(s_F/n_F)^2$ or $s_F^2/n_F^2 E$. With the latter form, except for constants which do not depend on pulse or meter shape, the ratio becomes

$$\frac{\left| \int_{-\infty}^{\infty} G_{IF}(\omega) S_I(\omega) \exp(i\omega t_0) d\omega \right|^2}{\int_{-\infty}^{\infty} |S_I(\omega)|^2 d\omega \int_{-\infty}^{\infty} |G_{IF}(\omega)|^2 d\omega}. \quad (52)$$

By the Schwartz inequality, this expression has the maximum value unity, and this is achieved

¹⁷ See, for example, E. C. Kemble, *Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1935), p. 36.

when

$$G_{IF}(\omega) = [S_I(\omega) \exp(i\omega t_0)]^* \\ = S_I(\omega)^* \exp(-i\omega t_0). \quad (53)$$

Equation (53) shows that except for a phase factor *the filter characteristic should be the conjugate of the Fourier transform of the pulse*, a condition we have quoted many times. The distinction between $S_I'(\omega)$ and $S_I(\omega)^*$ disappears for a symmetrical pulse, as then $S_I(\omega)$ is real. Otherwise one must match the filter to the conjugate of the Fourier transform rather than to the transform itself.

The theorem just given has already been derived, cf. North.¹ A different method of proof is usually employed, *viz.*, the calculus of variations. The possibility of using the Schwartz inequality, as we do here, has also been noted by Dr. Henry Wallman,¹⁸ and is contained as a special case of a more general theory due to Wiener. More important, in our opinion, than the mathematical nature of the proof, is the analysis and discussion given in the first paragraph of the present section, which shows that maximization of s_v/n_v also implies maximization of s_F/n_F . The usual proofs consider only the extremum of s_F/n_F , which mathematically does not *a priori* correspond to an extremum in s_v/n_v , though it is physically very reasonable that this should be the case. It is particularly to be emphasized that the maximum achievable value of the ratio (52) is quite independent of the shape of the noise spectrum and of the pulse $s_I(t)$. This means that with perfect matching the performance with visual presentation is independent of pulse shape, a result already quoted in Part I. In other words, *with our "simple" criterion*, the only thing that counts at match is just the input energy per pulse.

The phase factor $\exp(i\omega t_0)$ requires some discussion. It is determined by the time at which the signal is a maximum, though the magnitude of this maximum is unaffected by the choice of t_0 . At first sight it might seem that the time at which the signal reaches its maximum can be chosen at will by giving the phase factor $\exp(i\omega t_0)$ of the filter the proper value. However, this is

¹⁸ "Realizability of Filters" (M.I.T. Radiation Laboratory Report No. 637, Dec. 8, 1944).

not the case, for the filter must be capable of actual construction. On obvious physical grounds it is impossible to construct a filter which will give the response of a signal at the output before this signal even reaches the input. A filter which does not have this absurd characteristic we shall call "physically realizable." Mathematically, this can be shown equivalent to demanding that $G_{IF}(\omega)$ have no singularities in the lower half of the complex plane. Any polynomial structure of the type $\Pi_j(\omega - i\delta_j)^{-1}$, $\delta_j > 0$ is clearly physically realizable and corresponds to a series of cascaded, single-tuned circuits. The gaussian structure which we study rather explicitly later in the report is not physically realizable, but as shown by Wallman,¹⁸ this objection is rather academic. Namely, the gaussian behavior can be regarded as the limit of cascading a very large number of circuits. In the limit the time delay would have to be infinite, but the gaussian behavior, as far as frequency dependence is concerned, does not differ materially for practical purposes from the cascading of several circuits with only a finite time-lag. So although it is not possible to build a strictly gaussian system, one can find an actual system whose filtering action is very nearly equivalent to it, and, in fact, practical filters constructed by cascading will have an essentially gaussian response curve, to all intents and purposes.

(b) Aural Reception

We must now examine whether the Fourier transform criterion also applies to aural as well as to visual reception. Here the proof is not quite so immediate. First let us consider the square detector. As we are interested only in one audio-component, the effective signal-to-noise ratio, except for a constant factor A , is

$$(s_n/n_n)^2 = A \left(\int_{-\infty}^{\infty} s_F(t)^2 dt \right)^2 / \int_{-\infty}^{\infty} r_0(t)^2 dt, \quad (54)$$

where $r_0(t)$ is given by (12). The denominator of (54) is obtained directly from (17), and the numerator follows at once from (34). Since the Fourier transform of $s_F(t)$ is $S_I(\omega)G_{IF}(\omega)\sqrt{2}$, (vide Eq. (7b)), and that of $r_0(t)$ is proportional to $|G_{IF}(\omega)|^2$, by Plancherel's theorem (54) can be

written

$$(s_n/n_n)^2 = B \left(\int_{-\infty}^{\infty} |G_{IF}(\omega)|^2 |S_I(\omega)|^2 d\omega \right)^2 / \int_{-\infty}^{\infty} |G_{IF}(\omega)|^4 d\omega. \quad (55)$$

If we divide by $\int_{-\infty}^{\infty} |S_I(\omega)|^4 d\omega$, which is constant with respect to any given type of pulse, we have a ratio to which we can apply the Schwarz inequality. It follows immediately that for a maximum signal-to-noise ratio we must have

$$|S_I(\omega)|^2 = C |G_{IF}(\omega)|^2, \quad (56)$$

when C is an unimportant constant which we may set equal to unity without loss of generality.

Two remarks particularly should be made in connection with (56). One is that the question of the phase of the filter response does not arise, for the only quantity specified is $|G_{IF}(\omega)|$, and the phase shift can be an arbitrary function of frequency. In the visual case, on the other hand, the phase had to be the same as that of the conjugate of the Fourier transform of the signal, except for an arbitrary linear term determined by the time at which the signal was a maximum. Such a time obviously cannot enter in connection with audio detection, as here attention is focused on one particular component rather than on the maxima and minima obtained by superposing all the components.

The other remark is that with audio-reception the response at match is conditioned by shape rather than by just energy per pulse. If the latter alone were to enter, (55) should be divided by a factor $\int_{-\infty}^{\infty} |S_I(\omega)|^2 d\omega$, rather than $\int_{-\infty}^{\infty} |S_I(\omega)|^4 d\omega$. The resulting increased effectiveness of long pulses has already been discussed in Section V, Part I.

The proof that the Fourier transform criterion gives the best results in aural detection cannot be given in as clear-cut form for the linear as for the square-law detector. If we assume the signal is weak, so that we may use (39) and neglect all terms beyond $r_0^2(t)$ in (26), the ratio to be maximized is the same as for the square

detector and our previous proof will apply. Numerical examination of particular cases shows that the neglected terms do not disturb the location of the point of maximum efficiency. We may remark, moreover, that if it were not for the phenomenon of modulation suppression, the width of the IF filter would be a matter of indifference in aural detection. In actual practice, the true relationship is, so to speak, a compromise between (39) and (40), being much closer to the former than to the latter. Since in (40) the filter design does not enter, and since with (39) the Fourier transform criterion is valid, we can see qualitatively that in any intermediate case, use of this criterion will presumably bring the best results.

It can be shown, from the same sort of analysis, that the introduction of a video-filter at match gives no improvement of the signal to noise ratio. We will not give the details here, as the problem is amply covered elsewhere.¹⁹

IV. Explicit Calculation for a Gaussian Pulse and Filter

The general expression for the aural and visual signal-to-noise ratios have been obtained in Section II, Part II, but before we can arrive at complete numerical results on all aspects of the problem, we must specify the pulse $s_I(t)$ entering the IF and the response $G_{IF}(\omega)$ of the IF filter. Let the unfiltered envelope be

$$\sqrt{2}s_I(t) = S_0 \exp(-\omega_g^2 t^2/2), \quad \omega_g = 2\pi f_g, \quad (57)$$

where S_0 is the maximum amplitude of the disturbance, and ω_g is an angular frequency determining the width of the pulse. With the help of (5), the amplitude spectral distribution becomes

$$S_I(\omega) = (\pi/2)^{1/2} \omega_g^{-1} S_0 \exp(-\omega^2/2\omega_g^2). \quad (58)$$

Now for the (shifted) *amplitude* response of the IF filter we have

$$G_{IF}(\omega) = \exp(-\omega^2/2\omega_b^2), \quad \omega_b = 2\pi f_b, \quad (\omega = \omega' - \omega_0), \quad (59)$$

where ω_b similarly determines the filter width.

¹⁹ J. W. Lawson and G. E. Uhlenbeck, Radiation Laboratory, book 24.

In terms of the equivalent rectangular filter discussed in Section II, Part II, from Eq. (22) we find for the pulse and IF filter

$$\begin{aligned} \Delta f_{IF} &= (\Delta f_b) = \pi^{1/2} f_b; \\ \Delta f_I &= (\Delta f_g) = (1/\Delta T_g) = \pi^{1/2} f_g. \end{aligned} \quad (60)$$

With the aid of (7b) we see from (58) and (59) that the filtered wave becomes

$$\begin{aligned} \sqrt{2}s_F(t) &= (bS_0/\sqrt{2}\omega_g) \exp(-t^2 b^2/4) \\ &= \sqrt{2}s_F \exp(-t^2 b^2/4), \end{aligned} \quad (61a)$$

where

$$b \equiv \sqrt{2}\omega_b\omega_g/(\omega_g^2 + \omega_b^2)^{1/2}; \quad s_F \equiv bS_0/\sqrt{2}\omega_g. \quad (61b)$$

From (51) and (57) the input energy in a single gaussian pulse entering the IF is

$$E = \int_{-\infty}^{\infty} s_I(t)^2 dt = S_0^2 \pi^{1/2} / 2\omega_g, \quad (62)$$

and since the input power to the IF is (energy/pulse) $\times$ (no. of pulses), we have

$$P_{in} = E/T = S_0^2 \pi^{1/2} / 2\omega_g T, \quad (63)$$

where T is the period of the pulses. From (8) and (59) the mean square noise amplitude passed by the gaussian IF is

$$n_F^2 = \frac{W_0}{2\pi} \int_{-\infty}^{\infty} \exp(-\omega^2/\omega_b^2) d\omega = W_0 \omega_b / 2\pi^{1/2}, \quad (64)$$

so that from (61a, b) we find that

$$\begin{aligned} s_F(t)^2 / n_F^2 &= (s_F^2 / n_F^2) \exp(-t^2 b^2/2) \\ &= \frac{\pi^{1/2} b^2 S_0^2}{2W_0 \omega_b \omega_g^2} \exp(-t^2 b^2/2). \end{aligned} \quad (65)$$

Now let us define a quantity σ_0 by

$$\sigma_0 \equiv (s_F / n_F)_{\text{at match}}, \quad (66)$$

where σ_0 is a value of s_F/n_F , before rectification, which is obtained when a gaussian pulse, whose pulse repetition frequency is T_0^{-1} , enters an IF filter matched to the pulse, i.e., $\omega_b = \omega_g$. Let P_0 represent the input power to the IF under these circumstances. Then from (59) and (63) we obtain

$$\begin{aligned} \sigma_0^2 &= \pi^{1/2} S_0^2 / 2W_0 \omega_b, \\ \text{and} \quad P_0 &= S_0^2 \pi^{1/2} / 2\omega_b T_0 = W_0 \sigma_0^2 / T_0, \end{aligned} \quad (67)$$

which, applied to (65) gives us finally

$$s_F(t)^2/n_F^2 = (s_F^2/n_F^2) \exp(-t^2 b^2/2) \\ = \left(\frac{2\omega_g \omega_b}{\omega_g^2 + \omega_b^2} \right) \sigma_0^2 \left(\frac{P_{in}}{P_0} \frac{T}{T_0} \right) \exp(-t^2 b^2/2). \quad (68)$$

To examine the effects of variable pulse and filter widths on the possibly perceptible signal-to-noise ratios, we must then distinguish two cases:

Case I: Here the filter width ($\sim \omega_b$) remains constant, while the pulse width ($\sim \omega_g$) may have different values. Let

$$\tilde{\lambda} = \omega_g/\omega_b = f_g/f_b, \\ \text{and} \quad \therefore b = \sqrt{2}\omega_g(1+\tilde{\lambda}^2)^{-1/2}. \quad (69)$$

From (58) it follows that

$$(s_F/n_F)^2 = \frac{2\tilde{\lambda}}{1+\tilde{\lambda}^2} \sigma_0^2 \tilde{P} \tilde{T}, \quad (70)$$

where $\tilde{P} = P_{in}/P_0$, and $\tilde{T} = T/T_0$, a normalization made in terms of quantities experimentally determined when the pulse and filter are matched. In a similar fashion, we find for *Case II*, where the pulse width ($\sim \omega_g$) is kept fast and the filter width ($\sim \omega_b$) is varied

$$\lambda = \omega_b/\omega_g (= \tilde{\lambda}^{-1}), \\ \text{and} \quad \therefore b = \lambda \sqrt{2}\omega_g(1+\lambda^2)^{-1/2}, \quad (71)$$

and, like (70)

$$(s_F/n_F)^2 = \frac{2\lambda}{1+\lambda^2} \sigma_0^2 \tilde{P} \tilde{T}. \quad (72)$$

When (70) and (72) are used in calculations involving the linear rectifier we must be careful that $(s_F/n_F)^2$ does not exceed 9 or 10, in accordance with the approximation (42). See also Fig. 9.

(a). Visual Reception with a Quadratic or Linear Rectifier

For visual reception we may write at once from (46), (70), and (72) that

$$(s_F/n_F)_{quad} = \frac{2\Lambda}{1+\Lambda^2} \sigma_0^2 \tilde{P} \tilde{T}, \quad (73)$$

where $\Lambda = \lambda$ or $\tilde{\lambda}$, as the case may be. Now when

$(s_F/n_F)_{quad}$ represents the minimum detectable signal-to-noise ratio, $(s_F/n_F)/(s_F/n_F)_{at \ match}$ must equal unity, since the minimum perceivable signal is by definition a constant σ_0^2 , from Eq. (73). Then it is evident that

$$1 = (s_F/n_F)_{quad}/(s_F/n_F)_{at \ match} \Lambda = 1 = \frac{2\Lambda}{1+\Lambda^2} \tilde{P} \tilde{T}, \quad (74)$$

a somewhat special case of a more general result obtained by M. C. Wang and G. E. Uhlenbeck at (M.I.T.) Radiation Laboratory, and considered by them elsewhere.

For visual reception following a linear detector we may similarly write, from (47), (70), and (72),

$$(s_F/n_F)_{lin} = \rho_0 \left(\frac{2\Lambda}{1+\Lambda^2} \right) \tilde{P} \tilde{T} \sigma_0^2 \\ \times \left\{ 1 - 2\gamma \left(\frac{2\Lambda \sigma_0^2 \tilde{P} \tilde{T}}{1+\Lambda^2} \right)^{k-1} \right\}, \quad \begin{matrix} \rho_0 = 0.957, \\ \gamma = 0.0763, \\ k = 1.50. \end{matrix} \quad (75)$$

Analogous to (74) we have

$$1 = (s_F/n_F)_{lin}/(s_F/n_F)_{at \ match} \Lambda = 1 \\ = \frac{2\Lambda}{1+\Lambda^2} \tilde{P} \tilde{T} \left(\frac{1 - 2\gamma \{ 2\Lambda \sigma_0^2 \tilde{P} \tilde{T} / (1+\Lambda^2) \}^{k-1}}{1 - 2\gamma \sigma_0^{2k-2}} \right). \quad (76)$$

We see at once that if $(s_F/n_F)^2 \ll 1$, we may neglect the second terms of (75) and (76), so that both of these equations become identical with (73) and (74), for the square-law rectifier, on the slight additional assumption that we treat $\rho_0 = 1$.

(b). Aural Reception with a Quadratic and Linear Rectifier

For aural reception we must employ (48a, b) and (49a, b). The normalized correlation $r_0(t)$ is obtained for the gaussian case from (13), and is $r_0(t) = \exp(-t^2 \omega_b^2/4)$, whereupon ξ_0 , cf. Eq. (48b), becomes $(\pi/2\omega_b^2)^{1/2}$. Finally, with the aid of (48a, b), (65), (70), and (72), and following the procedure of Section (a) above we obtain

$$(s_a/n_a)_{quad}/(s_a/n_a)_{at \ match} \tilde{\lambda} = 1 \\ = [2/(1+\tilde{\lambda}^2)]^{1/2} \tilde{P}, \text{ Case I,} \quad (77a)$$

and for Case II

$$(s_a/n_a)_{quad}/(s_a/n_a)_{at \ match} \lambda = 1 \\ = [2\lambda/(1+\lambda^2)]^{1/2} \tilde{P}, \text{ Case II.} \quad (77b)$$

Equations (77a) and (77b) are discussed at some length in Section V, Part I. In a similar fashion we obtain from (49a, b), (65), (70), (77) and Section (a) the ratios corresponding to the linear detector:

$$\frac{(s_a/n_a)_{lin}}{(s_a/n_a)_{at\ match}} = 1 = \frac{[2/(1+\lambda^2)]^{\frac{1}{2}} \bar{P}(1-2\gamma k^{-1}\{[2\lambda/(1+\lambda^2)]\sigma_0^2 \bar{P} \bar{T}\}^{k-1})}{1-2\gamma k^{-1}\sigma_0^{2k-2}}, \text{ Case I,} \quad (78a)$$

and

$$\frac{(s_a/n_a)_{lin}}{(s_a/n_a)_{at\ match}} = 1 = \frac{[2\lambda/(1+\lambda^2)]^{\frac{1}{2}} \bar{P}(1-2\gamma k^{-1}\{[2\lambda/(1+\lambda^2)]\sigma_0^2 \bar{P} \bar{T}\}^{k-1})}{1-2\gamma k^{-1}\sigma_0^{2k-2}}, \text{ Case II.} \quad (78b)$$

These equations are also discussed in some detail in Sections V and VI, Part I, and are illustrated in Figs. 4 and 5.

The above aural signal-to-noise ratios have all been computed on the assumption that there is no gating of the noise before rectification. If a gate is in operation, usually before the IF or rectifier stages, the background audio-noise voltage is simply reduced by the square-root of the gating factor $G \equiv \tau_g/T$, where τ_g is the duration of the "on-" period, or transmitting interval of the gate, i.e., the time interval during which noise may enter the IF or rectifier. This is a simple multiplicative factor, provided the duration of the "on-" period is sufficiently long to permit the reception of consecutive signals from the particular target under scrutiny. A more thorough discussion of gating is given in Section VI, Part I. We may say, then, that for either type of rectifier the gated background audio-noise voltage (r.m.s.) becomes

$$n_{aG} = G^{\frac{1}{2}} n_a = (\tau_g/T)^{\frac{1}{2}} n_a. \quad (79)$$

When the effect of (79) is considered in deriving the results of (77a) and (77b), we observe that these latter are modified to

$$1 = \left(\frac{2}{1+\lambda^2} \right)^{\frac{1}{2}} \bar{P} \bar{T}^{\frac{1}{2}}, \text{ (I),}$$

and

$$1 = \left(\frac{2\lambda}{1+\lambda^2} \right)^{\frac{1}{2}} \bar{P} \bar{T}^{\frac{1}{2}}, \text{ (II).} \quad (80)$$

A similar modification takes place in (78a) and (78b), the right members of which are now multiplied by $\bar{T}^{\frac{1}{2}}$. Curves of (78a) and (78b) for gating are shown in Figs. 4 and 5, for the condition of match.

(c). Comparison of the Aural and Visual Behavior of the Linear and Quadratic Detectors

From the results of the preceding pages we are in a position to compare conveniently the visual performance of the two types of rectifier, and to make the comparison in the aural case, as well. Let us examine first, briefly, the ratio between the visual linear response and the corresponding one for the quadratic. From (73) and (75) we obtain at once

$$\begin{aligned} (s_v/n_v)_{lin}/(s_v/n_v)_{quad} \\ = \rho_0 \left\{ 1 - 2\gamma \left[\frac{2\Lambda}{1+\Lambda^2} \sigma_0^2 \bar{P} \bar{T} \right]^{k-1} \right\} \end{aligned} \quad (81)$$

which holds equally well for either Case I or Case II. At match $\bar{P} = \bar{T} = 1$ and $\Lambda = 1$, and we have

$$\begin{aligned} [(s_v/n_v)_{lin}/(s_v/n_v)_{quad}]_{match} \\ = 0.957(1 - 0.1526\sigma_0), \quad \sigma_0 \leq 10^{\frac{1}{2}}. \end{aligned} \quad (82)$$

For aural reception we obtain expressions analogous to (81) and (82) by the same process used in arriving at (77-78), namely

$$\begin{aligned} (s_a/n_a)_{lin}/(s_a/n_a)_{quad} \\ = \mu_0 \left(1 - 2\gamma k^{-1} \left\{ \frac{2\Lambda}{1+\Lambda^2} \sigma_0^2 \bar{P} \bar{T} \right\}^{k-1} \right), \end{aligned} \quad (83)$$

where

$$\mu_0 = \left\{ \sum_{n=0}^{\infty} \frac{[(2n)!/2^{2n}n!(n+1)!] \bar{P}^{\frac{1}{2}}}{(n+1)^{\frac{1}{2}}} \right\}^{-1} = 0.973,$$

and where either f_b or f_s may be constant. Equation (83) becomes particularly simple at match ($\Lambda = \bar{P} = \bar{T} = 1$), where $k = 1.50$, yielding

$$\begin{aligned} \{(s_a/n_a)_{lin}/(s_a/n_a)_{quad}\}_{at\ match} \\ = 0.973(1 - 0.1246\sigma_0), \quad \sigma_0 \leq 10^{\frac{1}{2}}, \end{aligned} \quad (84)$$

which is not very different from (82). It might be pointed out that (82) is independent of pulse and filter shape, the only requirement being that the two are properly matched. This is very nearly true of (84), though not strictly so, as the factor k^{-1} in (83) appears as a consequence of the gaussian structure of the filter and pulse. However, for all practical purposes we may consider the aural comparison also independent of pulse and filter shape, at match. Even when a condition of match does not obtain, if $(s_F/n_F)^2 \ll 1$, i.e., $2\Delta\sigma_0^2 P T / (1+\Delta)^2 \ll 1$, which is the usual, practical situation, the performance of the two rectifiers is identical. Figure 2 demonstrates this, and shows how the behavior of the linear detector departs from that of quadratic. This departure in the customary region of operation does not exceed 1.5 db. Note that the ratios (83) and (84) hold equally well for gated noise, in the aural case.

V. General Theory for the Meter

Instead of listening to the fundamental (or a harmonic) of the PRF against the noise background, passed by a narrow audio-filter, we may rectify this component and its surrounding noise, and pass the mixture into a meter with a very narrow low pass filter, say from $1/\tau_0$ to 1 cycle wide. Our problem is to obtain an expression for the signal-to-noise ratio as recorded by the meter, i.e., to determine s_m/n_m and to examine how this ratio depends on such factors as audio- and meter filter widths, signal to noise in the audio-filter, etc.

Let us consider first the noise entering the meter. This results from the rectification of the noise passed by the narrow audio-filter centered about the harmonic of the PRF under examination. Because this filter is quite narrow, say 10–100 cycles, and conceivably less, we cannot always assume as we did for the case of aural detection, that the resulting very low frequency continuum, $W_{LF}(\omega)_a$, has a uniform spectral density, and can be replaced by $W_{LF}(0)_a$ in our derivation. We then write, following (21a), that the mean noise power in the meter is

$$n_m^2 = \frac{1}{2\pi} \int_0^\infty |G_m(\omega)|^2 W_{LF}(\omega)_a d\omega, \quad (85)$$

where $G_m(\omega)$ is the amplitude response of the low

pass meter filter. Although the low frequency noise is no longer gaussian after rectification by the second detector, the audio-filter is assumed sufficiently narrow so that the noise passed by it may be considered random, with its gaussian characteristics restored.²⁰ Then, to determine the low frequency spectrum after rectification we need only parallel the analysis of Section II, Part II, except that the signal, after being passed by the audio-filter, is now periodic, rather than highly pulsed. From (26) applied to the narrow portion of the *audio*-spectrum, instead of to the IF, we obtain for a linear rectifier in the meter circuit

$$W_{LF}(\omega)_a = \frac{n_a^2 \beta_0^2}{2\pi} \sum_{n=0}^{\infty} \left\{ \frac{(2n)!}{2^{2n} n! (n+1)!} \right\}^2 \times \int_0^\infty r_0(t)_a^{2n+2} \cos \omega t dt, \quad (86)$$

where β_0 is the transconductance of the rectifier, and $r_0(t)_a$ is the normalized correlation function of the random noise passed by the audio-filter, which may easily be obtained from (12) and (13) on replacing $G_{IF}(\omega)$ by $G_a(\omega)$. Since the audio-filter is narrow compared to the PRF harmonic, we may safely extend the limits on the integration to $-\infty$ in determining $r_0(t)_a$, and for the mean audio-noise power we find

$$n_a^2 = \frac{W_{LF}(0)}{2\pi} \int_{-\infty}^{\infty} |G_a(\omega)|^2 d\omega, \quad (87)$$

from (21b). As before, the argument of $G_a(\omega)$ is the frequency relative to an origin at the carrier, now the PRF or one of its harmonics. $W_{LF}(0)$, of course, depends on the characteristic of the second detector used after the IF. If the noise is gated, we may replace n_a by n_{a0} in subsequent formulae, remembering that the two forms are related by Eq. (79).

In case the rectification is quadratic rather than linear, we may write in place of (86), in virtue of (17) and (19),

$$W_{LF}(\omega)_a = 4\eta_0^2 n_a^4 \int_0^\infty r_0(t)_a^2 \cos \omega t dt, \quad (88)$$

with η_0 the detector constant.

²⁰ See the first part of section 4.3, of Rice's article (reference 8).

The signal-to-noise ratio in the meter is obtained at once by analogy with (38) and (42) for the linear rectifier, and is

$$s_m/n_m = \frac{\beta_0}{2(2\pi)^{\frac{1}{2}}} \left(\frac{n_{a(G)}}{n_m} \right) \{ (s_a/n_{a(G)})^2 - 2\gamma (s_a/n_{a(G)})^{2k} \}, \quad s_a/n_{a(G)} < 10^{\frac{1}{2}}. \quad (89)$$

For the quadratic detector, Eq. (89) is modified by virtue of (29) to

$$s_m/n_m = \eta_0 s_a^2 / n_m. \quad (90)$$

The expressions for $s_a/n_{a(G)}$ depend on the circuit preceding the meter. The proper forms are given in Section II, (e), Part II. With the above results we are now ready to proceed to specific problems.

VI. An Explicit Treatment for the Meter

Let us determine s_m/n_m for a meter following a linear rectifier. We assume further that the audio- and meter filters are essentially gaussian and have the responses

$$\begin{aligned} G_a(\omega) &= \exp(-\omega^2/2\omega_a^2), \\ G_m(\omega) &= \exp(-\omega^2/2\omega_m^2), \end{aligned} \quad (91)$$

where ω_a and ω_m are analogous to ω_b . The correlation $r_0(t)_a$ becomes with the help of (91)

$$r_0(t)_a = \exp(-t^2\omega_a^2/4). \quad (92)$$

With this result we may write for the low frequency spectrum after rectification, from (86),

$$\begin{aligned} W_{I, F}(\omega)_a &= \frac{n_a^2 \beta_0^2}{2\omega_a(2\pi)^{\frac{1}{2}}} \\ &\times \sum_{n=0}^{\infty} \left\{ \frac{(2n)!}{2^{2n} n! (n+1)! (n+1)!} \right\}^2 \\ &\times \exp[-\omega^2/2\omega_a^2(n+1)]. \end{aligned} \quad (93)$$

Now we find from (85) that the mean noise power is

$$\begin{aligned} n_m^2 &= \frac{n_a^2 \beta_0^2}{8\pi\sqrt{2}} \sum_{n=0}^{\infty} \left\{ \frac{(2n)!}{2^{2n} n! (n+1)! (n+1)!} \right\}^2 \\ &\times \frac{(\omega_m/\omega_a)}{[1 + (\omega_m/\omega_a)^2 \frac{1}{2} (n+1)^{-1}]^{\frac{1}{2}}}. \end{aligned} \quad (94)$$

Applying this directly to (89) we obtain for the signal-to-noise ratio in the meter

$$s_m/n_m = 2^{\frac{1}{2}} (\sum_n)^{-\frac{1}{2}} \times \{ (s_a/n_{a(G)})^2 - 2\gamma (s_a/n_{a(G)})^{2k} \}, \quad (95)$$

where $\sum_n$ is the series in (94). Our result (95) may be simplified if we disregard the terms in $(\omega_m/\omega_a)^2 [1/2(n+1)]$, $n \geq 1$, which appear in $\sum_n$, and which for our problem are quite small compared to the term in $n=0$. This approximation becomes progressively better the smaller the ratio ω_m/ω_a . Then the summation takes the value

$$\sum_n = \frac{1.059(\omega_m/\omega_a)}{[1 + \frac{1}{2}(\omega_m/\omega_a)^2]^{\frac{1}{2}}} = \frac{1.059(\Delta f_m/\Delta f_a)}{[1 + \frac{1}{2}(\Delta f_m/\Delta f_a)^2]^{\frac{1}{2}}}, \quad (96)$$

where Δf_m and Δf_a are the widths of the equivalent rectangular filters, *vide* Eq. (22). We have then for gated noise

$$\begin{aligned} \frac{s_m}{n_m} &= 1.156 \frac{\{1 + \frac{1}{2}(\Delta f_m/\Delta f_a)^2\}^{\frac{1}{2}}}{(\Delta f_m/\Delta f_a)^{\frac{1}{2}}} \\ &\times \left\{ (s_a/n_{a(G)})^2 - 2\gamma \left(\frac{s_a}{n_{a(G)}} \right)^{2k} \right\}, \quad s_a/n_{a(G)} \leq 10^{\frac{1}{2}}. \end{aligned} \quad (97)$$

With this result we are at once able to compare, for a given set of operating conditions, the relative merits of the meter, the "ear-scope," or carphone detection, and the A-scope, provided we know in each case what the physically minimum values of s_m/n_m , $s_a/n_{a(G)}$, s_v/n_v are. Once these are known, the comparison follows immediately, as they are all functions of the single quantity s_F/n_F . (See Section VII, Part I.)

We are particularly interested in determining, for a given apparatus, how wide the audio-filter, Δf , must be in the "earscope" method, as compared with the filters used in connection with periodic meter detection. We will assume that the minimum detectable signal-to-noise ratios are related in either case by

$$s_m/n_m = A [s_a/n_{a(G)}]_{\text{earscope}}, \quad (98)$$

where the constant A is determined experimentally. One condition of comparison is then, from (98)

$$[s_a/n_{a(G)}]_{\text{earscope}} = a_0/A, \quad (99)$$

where now α_0 is the value of the minimum detectable signal-to-noise ratio when a periodic metering scheme is used, into which $s_a/n_{a(G)} = \alpha$ has been introduced. The quantity α may be defined by the following relation

$$[s_a/n_{a(G)}]_{\text{earscope}} \equiv \alpha(\Delta f_a/\Delta f_s)^{\frac{1}{2}} \\ = \left[\frac{s_a}{n_{a(G)}} \right] \left(\frac{\Delta f_a}{\Delta f_s} \right)^{\frac{1}{2}}, \quad (100)$$

which is an obvious consequence of the fact that the noise power passed by the audio-filter is proportional to its width. Using (97)–(100), we find first that

$$\alpha = \alpha_0(\Delta f_s/\Delta f_m)^{\frac{1}{2}}/(\Delta f_a/\Delta f_m)^{\frac{1}{2}}. \quad (101)$$

Letting $y \equiv \Delta f_s/\Delta f_m$ and $x \equiv \Delta f_a/\Delta f_m$, we have with the help of (97) and (101)

$$\alpha_0 = 1.156(1 + \frac{1}{2}x^{-2})^{\frac{1}{2}} \left\{ \left[\frac{\alpha_0}{A} (y/x)^{\frac{1}{2}} \right]^2 - 2\gamma \left[\frac{\alpha_0}{A} (y/x)^{\frac{1}{2}} \right] \right\}, \quad \gamma = 0.0763, \quad (102)$$

which reduces simply to

$$1 = (y/x)^{\frac{1}{2}}(2x^2 + 1)(\alpha_0/A^2 1.059)^{\frac{1}{2}} \\ \times \{1 - 2\gamma\alpha_0(y/x)^{\frac{1}{2}}/A\}. \quad (103)$$

Let

$$a_0 \equiv \alpha_0/A^2(1.059)^{\frac{1}{2}} = 0.971\alpha_0/A^2; \\ b_0 \equiv 2\gamma\alpha_0/A = 0.1526\alpha_0/A; \quad z \equiv y/x. \quad (104)$$

Then (103) becomes

$$(za_0)^{\frac{1}{2}}(2x^2 + 1)(1 - b_0z^{\frac{1}{2}})^{\frac{1}{2}} = 1, \quad (105)$$

subject to the requirement that z is always chosen so that $(1 - b_0z^{\frac{1}{2}})za_0 < 1$. If the detector just preceding the meter is quadratic, instead of linear, Eq. (97) may be written, from (85), (88), (90):

$$\frac{s_m}{n_m} = 2^{\frac{1}{2}} \frac{\{1 + \frac{1}{2}(\Delta f_m/\Delta f_s)^2\}^{\frac{1}{2}}}{(\Delta f_m/\Delta f_s)^{\frac{1}{2}}} \\ \times \left[\frac{s_a}{n_{a(G)}} \right]^2; \quad 2^{\frac{1}{2}} = 1.189. \quad (106)$$

This reduces simply to

$$z = \frac{A^2}{\alpha_0} (2x^2 + 1)^{-1}$$

or

$$y = A^2 x (2x^2 + 1)^{-1} / \alpha_0. \quad (107)$$

If we plot y against x we have the required relationship between Δf_s and Δf_a . Curves of (105) and (107) are shown in Fig. 3 for different values of the minimum detectable signal α_0 , $A = 1$. The case $A = 1$ is particularly important, as it enables us to compare the periodic and aperiodic metering schemes, where now $\Delta f_s = (\Delta f_m)_{\text{periodic}}$. See Section IV, Part 1, and Figs. 3, 8 for a more detailed discussion.

We would like to express our sincere thanks for helpful comments and criticisms to Drs. E. R. Brill, W. H. Jordan, D. O. North, D. Taylor, and H. Wallman.

The Permeability of Different Rubbers to Gases and Its Relation to Diffusivity and Solubility*

G. J. VAN AMERONGEN
Rubber-Foundation, Delft, Holland

(Received June 6, 1946)

As the permeability of rubber-like substances to gases stands in relation to the solubility and rate of diffusion of the gases in those materials, these individual values should be known. The permeability of a membrane was measured manometrically and the diffusivity was derived from the time-lag of the permeation. The solubility of the gas was computed from the permeability and the diffusivity, in addition to which the solubility was also found by direct measurement. In this way eight different gases were tested with nine elastomers at different temperatures. It appeared that the permeability of a membrane to a given gas is not affected by the presence of a second gas. The differences in permeability of different elastomers to a given gas are caused mainly by differences in rate of diffusion and only in a very minor degree to differences in solubility. The differences in permeability of the same elastomer to different gases are caused not only by differences in rate of diffusion but also by differences in solubility. A linear relationship is found between the logarithm of the solubilities of different gases in natural rubber and their critical temperatures, so the higher the

critical temperature of a gas, the better does it dissolve. The presence of polar groups in an elastomer reduces the solubility of non-polar gases and increases the solubility of polar gases in the elastomer. The various rubbers behave towards gases like organic liquids. The activation energy of the diffusion and the heat of solution were calculated from the temperature function of the diffusivity and the solubility. As the diameter of the molecule of the gas increases, the rate of diffusion decreases, while the activation energy of the diffusion becomes greater. The presence of polar groups and methyl groups in elastomers causes low rate of diffusion, which involves a great activation energy of diffusion. It is presumed that the activation energy of the diffusion is required to separate the rubber molecules for the displacement of the gas molecules. The attempt to elucidate the constant D_0 in the equation $D = D_0 \exp(-E/RT)$ —which proves to be a function of the activation energy of the diffusion E —by reference to one of the formulas published in the literature failed. An empirical formula was drawn up, by which D_0 is directly related to the activation energy E .

1. INTRODUCTION

AS far back as 1831 Mitchell¹ discovered that a rubber membrane is permeable to gases and that the rate of permeation is different for different gases. With the advance of synthetic rubbers in recent years it furthermore became evident that there was a marked difference in the behavior of these various types of rubber in the presence of a given gas. While this divergence in the behavior of different gases and kinds of rubber is primarily of great technical importance (one has only to think of balloons and inner tubes), it also offers interesting aspects from the theoretical point of view.

All the experimental data up to date² go to show that, with rubber, the process of permeation is one of solution of the gas molecules on one side of the membrane followed by diffusion through the rubber to the other side, where the gas

evaporates. Another fact which has emerged is that the quantity of dissolved gas in the rubber is proportional to the pressure, following Henry's law, while Fick's law applies to the actual process of diffusion. On this assumption, the quantity of permeating gas q is determined by³

$$q = DhA[(p_1 - p_2)/d]t, \quad (1)$$

where

D = the diffusivity,

h = the solubility,

A = the area of the membrane,

d = the thickness of the membrane,

p_1 and p_2 = the pressures on both sides of the membrane, and
 t = the time.

If unit values are taken for the difference in pressure $p_1 - p_2$, the area A , the thickness d , and the time t , then

$$Q = Dh, \quad (2)$$

where Q stands for the permeability.

The units which will be used can be seen from this formula. According to *Bunsen*, the solu-

* Communication No. 55 of the Rubber-Foundation, Delft, Holland.

¹ J. V. Mitchell, *J. Roy. Inst.* 2, 101, 307 (1831).

² Cf. H. A. Daynes, *Trans. Inst. Rubber Ind.* 3, 428 (1937).

³ S. von Wroblewski, *Ann. d. physik. Chem.* 8, 29 (1879).

bility k is expressed by the number of cubic cm of gas at 0°C and 76 cm Hg which dissolves per cu. cm of rubber and 1 atmos. gas pressure. The diffusivity D is expressed in sq. cm per second and the permeability Q by the number of cu. cm of gas at 0°C and 76 cm Hg which penetrates 1 sq. cm of the surface of a piece of material 1 cm thick per second when the outside difference in pressure of the gas amounts to 1 atmosphere. The dimensions of Q are

$$[P_t^{-1}p^{-1}] \text{ or } [P_m^{-1}t].$$

It is necessary to add that the above formulas apply only if the rate of solution and evaporation is very great in comparison to the rate of diffusion, in other words, if the concentration of the gas in the rubber interface remains constant and equal to the concentration of equilibrium (according to Henry's law), irrespective of the rate of diffusion. For instance, permeation is a more complicated phenomenon with metals, when the rate of solution or evaporation of the gas often governs the rate of permeation. This will be evident chiefly by the fact that the rate of permeation is no longer in inverse ratio to the thickness of the membrane. It has been experimentally proved, however, that formula (1) holds good for rubber, so with that material there is no reason to assume that the rate of solution has much influence upon the rate of permeation.

Our purpose now was to find out why a given kind of rubber possesses a distinctive permeability. To do this it will not suffice to make permeability measurements alone; it will be necessary to measure separately the factors, solubility and rate of diffusion of the gas in the elastomer, which, together, determine the permeability. It is best, moreover, to determine these two factors in dependence upon the temperature, since the heat of solution and the activation energy of the diffusion can then be calculated, by which means better insight into the whole process can be obtained.

2. METHOD FOR THE DETERMINATION OF THE PERMEABILITY

The apparatus with the aid of which the permeability and rate of diffusion were measured works on the same principles as that designed by

De Boer and Fast⁴ and by Barrer,⁵ the characteristic feature of which is a permeability cell with a pressure gauge, as represented in Fig. 1.

A gas cell is made up of two parts of steel, A and B , containing gas chambers separated from each other by a rubber membrane. The gas is introduced into the A chamber, the volume of which is approximately 75 cu. cm, while the B chamber, which is connected to a pressure gauge, is exhausted to approximately 0.1 mm Hg pressure. To prevent the membrane from sagging, it is supported on fine gauze netting which, in turn, rests on a perforated iron table.

The flanges, when warmed, were given a thin coat of an asphaltic bitumen mixture so that, once they had been screwed down, the apparatus was completely gas-tight.

A McLeod pressure gauge, with a range up to 2.5 mm Hg pressure and recording with an accuracy of 0.005 mm Hg pressure, was used to measure the pressure in the vacuum chamber. By noting for some time the increase in pressure in the vacuum chamber due to the permeability of the rubber, the permeability Q can be found

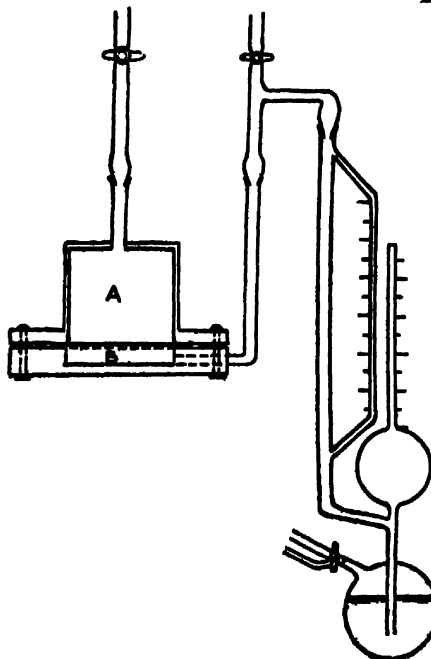


FIG. 1. Apparatus for the measurement of the permeation and diffusion of gases in rubber.

⁴ J. H. De Boer and J. D. Fast, *Rec. trav. chim.* **57**, 317 (1938).

⁵ R. M. Barrer, *Trans. Faraday Soc.* **35**, 628 (1939).

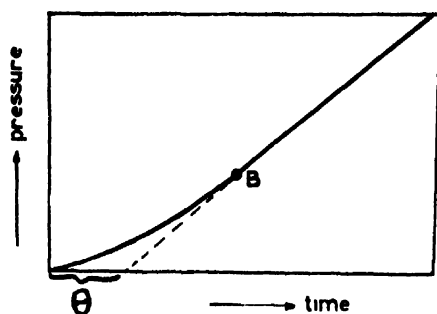


FIG. 2. Relation between increase in pressure and the time.

with the aid of the equation

$$Q = \frac{dp}{dt} \times \frac{V}{p} \times \frac{d}{A} \times \frac{273}{T}, \quad (3)$$

where

dp = increase in the pressure in B in time dt ,
 p = gas pressure in the gas chamber A,
 V = volume of the apparatus on the vacuum side,
 d = thickness of the membrane,
 A = area of the membrane, and
 T = average temperature of the vacuum chamber.

The surface area A of the membrane was always 30.2 sq. cm. The thickness of the membrane was determined with a Schopper thixometer to within an accuracy of 0.005 to 0.01 mm. The gas pressure p , which should be constant during a test, was usually 760 mm Hg. So that the measurements could be made at constant temperature, the gas cell was placed in a water thermostat. The membranes were made as described on a former occasion.⁶

This apparatus has much to recommend it for permeability determinations, which can often be carried out very quickly and with considerable recording accuracy; the consumption of gas per determination is negligible. The determination of Q is accurate to within 5 percent.

3. METHOD FOR THE DETERMINATION OF THE RATE OF DIFFUSION

Using the foregoing permeability apparatus, the rate of diffusion can be measured, in combination with the permeability determinations, by a method evolved by Daynes⁷ and Barrer.⁸ If a gas is brought into contact with one side of

a gas-free rubber membrane, it will take some time before the gas has reached the other side, which can be observed experimentally if the pressure of the vacuum (pressure gauge) side is plotted against the time, as in Fig. 2.

It takes some time before there is constant increase in pressure, from which the permeability is computed. Daynes succeeded in finding an equation which correlates the time-lag extrapolated to the zero axis θ , after which permeation begins, the diffusivity D and the thickness d of the membrane, starting from the diffusion equation

$$\delta C / \delta t = D \delta^2 C / \delta x^2. \quad (4)$$

The boundary conditions for the problem set are that:

1. for $x=0$ and all t $C=C_0$,
2. for $x=d$ and all t $C=0$,
3. for $t=0$ and all x $C=0$, and
4. A stationary condition prevails at $t=\infty$, so $\delta C / \delta t = 0$.

From this an equation can be obtained which gives the concentration C of the gas in the membrane as a function of time t and place x ;

$$C = C_0 \left(1 - \frac{x}{d} \right) - \frac{2C_0}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin \frac{n\pi x}{d} \exp \left[- \left(\frac{n\pi}{d} \right)^2 D t \right]. \quad (5)$$

The amount of material which at the place $x=d$ in time t passes through the area A is

$$q = AD \int_0^t - (dC/dx)_{x=d} dt. \quad (6)$$

After determination of dC/dx from (5), substitution in (6) and integration, we get

$$q = AD \left\{ \frac{C_0 t}{d} - \frac{C_0 d}{6D} \sum_{n=1}^{\infty} 1 - \exp \left[- \left(\frac{n\pi}{d} \right)^2 D t \right] \right\}. \quad (7)$$

To the stationary stage applies

$$q = AD \frac{C_0 d}{d}. \quad (8)$$

If $t = \theta + t_1$, then by combination of (7) and (8)

$$D = d^2 / 6\theta. \quad (9)$$

Thus Eq. (9) offers a simple means of calculating the diffusivity D from the thickness d of the

⁶ G. J. van Amerongen, Rev. Gén. du Caoutchouc 21, 50 (1944).

⁷ H. A. Daynes, Proc. Roy. Soc. A97, 286 (1920).

membrane and the time-lag θ of the permeability determination.

It is very important to know at what point the stationary condition begins.

In this point (B of Fig. 2) the term

$$\sum_{n=1}^{\infty} \exp \left[- \left(\frac{n\pi}{d} \right)^2 D t \right]$$

in Eq. (7) approaches zero,

$$\sum_{n=1}^{\infty} \exp \left[- \left(\frac{n\pi}{d} \right)^2 D t \right] = \sum_{n=1}^{\infty} \exp \left[- \frac{n^2 \pi^2 t}{6\theta} \right]$$

because $D = d^2/6\theta$,

$$\sum_{n=1}^{\infty} \exp \left[- \frac{n^2 \pi^2 t}{6\theta} \right] = \exp \left[- \frac{\pi^2 t}{6\theta} \right] + \exp \left[- \frac{4\pi^2 t}{6\theta} \right] + \exp \left[- \frac{9\pi^2 t}{6\theta} \right] + \dots \quad (10)$$

If various values are put into (10) for t , it will be found that this term can practically be neglected after a time which is 2.5 to 3 times that of the time-lag, and indeed it is usually from this point that extrapolation has taken place to the zero axis.

In using Eq. (9) it has been assumed that the pressure in chamber B of Fig. 1, which at most reached 2 to 3 mm Hg, may be disregarded in comparison with the pressure in A , which usually amounted to 760 mm Hg. The determination of D is accurate to 5-10 percent.

The rate of diffusion through some materials is so slow that, as a result, the determinations sometimes take hours to carry out. For example, the time-lag of the permeation of N_2 through a membrane of Oppanol 0.20 mm thick is 40 minutes at 17°C , so that 100 to 120 minutes elapse before the stationary condition sets in.

It need hardly be said that the membrane has to be de-gassed after each test by evacuation for a period which is 3 to 4 times the time-lag.

An alternative to this method is to find the diffusivity as the quotient of the permeability and solubility, provided the solubility has been measured by some independent method, as described in the following section.

4. METHOD FOR THE DETERMINATION OF THE SOLUBILITY

It is possible to calculate indirectly the solubility of a gas in rubber from Eq. (2), which

shows it to be the ratio of the permeability Q to the diffusivity D .

This calculation requires, of course, the independent determination of Q and D without the use of Eq. (2). A direct measurement of the solubility may be made with the apparatus shown in Fig. 3.

A flask A , filled with fine shippings of rubber, is connected to a gas volume meter BC sealed with mercury. After chambers A and B have been exhausted, cock K is closed and chamber B filled with the gas under test. The gas is raised to exactly one atmosphere pressure by means of a movable mercury tube C . Cock K is then opened and the gas streams from chamber B to A . After waiting until equilibrium is attained, constantly adjusting tube C in order to maintain pressure at 1 atmosphere, the gas volume is read off. This volume of gas, less the non-rubber volume of flask A , gives us the quantity of gas absorbed by the rubber. Determination can be made at various temperatures by suspending A up to cock K in a water thermostat.

With most gases the test is accurate to 5 percent. The solubility of CO_2 was determined to 1 to 2 percent accuracy, as this was the most suitable of the gases for the apparatus.

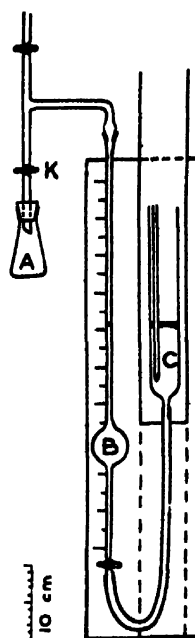


FIG. 3. Equipment for measuring the solubility of gases in rubber.

TABLE I. The permeability Q (in 10^{-8} cm³ sec.⁻¹ atm.⁻¹) and the diffusivity D (in 10^{-7} cm² sec.⁻¹) of gas in various elastomers.

t in °C	H ₂			O ₂			N ₂			CO ₂			CH ₄		He
	Q	$D = \frac{Q}{h}$	$D = \frac{Q}{h}$	Q	$D = \frac{Q}{h}$	$D = \frac{Q}{h}$	Q	$D = \frac{Q}{h}$	$D = \frac{Q}{h}$	Q	$D = \frac{Q}{h}$	$D = \frac{Q}{h}$	Q	$D = \frac{Q}{h}$	Q
Natural rubber															
17°	28	79	76	12	12.5	12	4.1	8.0	8.3	72	6.7	7.1	12.5		16.5
25°	39	105	100	18	17.5	18	6.6	11.5	12.5	102	10.5	11	22	8.8	23
35°	58.5	140	140	28.5	27	28	11.0	20	21	145	17	18.5	36		33
43°	77	185	180	39	36	39	16	28	29	185	25	26	50	21	44
50°	97	220	230	49.5	49	49	22.5	37	40	220	32	34	64		
Buna S															
17°	22.5	80		9.0	9.6		3.0	7.2		71	6.8		10.5		13
25°	30.5	100	85	13	14	14	4.8	10	10	94	10	11	16		17.5
35°	44	135		20	20		7.8	14.5		130	15.5		26		25
43°	59.5	165	150	27.5	28	30	11.5	21	23	165	23	24	35		33
50°	74	200		34.5	34		14.5	28		195	29		43		42
Perbunan															
17°	7.6	31		2.0	2.4		0.50	1.45		15	1.0		1.3		5.8
25°	11.5	42	41	3.2	3.6	4.0	0.89	2.3	2.5	23	1.7	1.8	2.4		8.7
35°	17.5	64		5.3	6.3		1.65	4.1		37	3.1		4.7		12.5
43°	25.5	86	82	7.7	9.1	9.6	2.5	6.2	6.3	52	4.7	5.3	7.0		16.5
50°	31.5	110		10.5	13		3.7	8.6		66	7.0		10.1		21
Neoprene G															
17°	6.8	29		1.75	2.5		0.53	1.55		12.4	1.3		1.4		
25°	10.3	38	36	3.0	3.8	4.0	0.89	2.4	2.5	19.5	2.3	2.4	2.5		
35°	16	56		5.1	6.2		1.65	4.4		31	4.2		4.8		
43°	23	74	72	7.7	10.0	10.0	2.55	7.2	6.7	43.5	6.8	6.8	7.2		
50°	28.5	94		10.1	13		3.55	9.1		56.5	9.1		9.8		
Oppanol B 200															
17°	3.0	10		0.54	0.50		0.11	0.27		2.3	0.32		0.2		3.8
25°	4.9	14	14	0.90	0.78	0.84	0.22	0.43	0.42	3.8	0.54	0.55	0.55		5.6
35°	8.3	21		1.6	1.55		0.44	0.84		6.6	1.05		1.1		8.5
43°	12	31	30	2.6	2.4	2.5	0.78	1.5	1.4	10.0	1.8	1.8	1.9		12
50°	16.5	38		3.7	3.4		1.15	2.1		14	2.5		2.9		15.5
Butadiene rubber															
17°	23	75		10.1	11		3.4	8.1		80	7.6				
25°	32	96		14.5	15		4.9	11		105	10.5				
35°	46	125		21	22		7.8	16		140	16				
43°	60	160		28	30		11	22		175	22				
50°	77	180		36	37		14.5	29		200	28				
Methyl rubber															
17°	9.0	27		0.93	0.88		0.20	0.46		3.0	0.36				8.1
25°	13	39		1.6	1.4		0.36	0.79		5.7	0.63		0.60		11
35°	20	61		3.1	2.5		0.78	1.7		10.5	1.3		1.45		16
43°	29	80		5.0	4.1		1.3	2.8		17	2.2		2.7		22
50°	38	105		7.1	6.1		2.2	4.1		24	3.6		4.4		27
Mipolam MP															
17°	2.9	15		0.41	0.90					2.3	0.23				
25°	4.4	20		0.70	1.3		0.2			4.0	0.44				
35°	7.3	30		1.4	2.4		0.4			7.6	0.83				
43°	10.5	39		2.1	3.9		0.7			11.5	1.4				
50°	14	50		3.2	5.2					15	1.9				
Thiokol B															
17°	0.71	7.8		0.11						1.3	0.43				
25°	1.2	10.5		0.22						2.4	0.81				
35°	2.2	18		0.49						4.8	1.7				
43°	3.3	26		0.85						7.7	2.8				
50°	4.6	35		1.3						11.0	4.0				

5. EXPERIMENTAL TESTS

The rubbers used for the determinations were natural rubber, Buna S, Perbunan, methyl rubber (polydimethylbutadiene) and butadiene rubber (polybutadiene) vulcanized for 30 minutes at 142°C in the proportions of 100 parts of rubber, 1.5 parts of sulphur, 5 parts of zinc oxide, 2 parts of stearic acid, 1.8 parts of Vulkacite AZ, 1 part of phenyl- β -naphthylamine. Without any additions, Neoprene G was vulcanized for 15 minutes at 142°C, while Thiokol B was vulcanized with 10 parts of zinc oxide and

1 part of stearic acid for 25 minutes at 142°C. Thus any further mention of a particular kind of rubber *always* connotes a vulcanized one. This does not apply to Oppanol B 200 (polyisobutene) and Mipolam MP (vinylchloride-methylacrylate copolymer), which cannot be vulcanized and therefore required no admixtures, though, as a commercial product, Mipolam MP contains 27 percent of softener.

The principal tests were preceded by a few H₂ permeability determinations with natural rubber in dependence upon pressure and thickness in

TABLE II. The solubility h of gas in various elastomers.

t in °C	H ₂		O ₂		N ₂		CO ₂		CH ₄	NH ₃	SO ₂
	Direct	$h = \frac{Q}{D}$	Direct	$h = \frac{Q}{D}$	Direct	$h = \frac{Q}{D}$	Direct	$h = \frac{Q}{D}$	Direct	Direct	Direct
<i>Natural rubber</i>											
17°	0.037	0.036	0.099	0.100	0.049	0.054	1.01	1.08	0.26		
25°	0.039	0.038	0.099	0.102	0.052	0.056	0.90	0.97	0.25	6.9	23.6
35°	0.041	0.040	0.100	0.104	0.053	0.058	0.78	0.83	0.25		
43°	0.042	0.042	0.100	0.105	0.055	0.059	0.70	0.74	0.24	4.1	11.6
50°	0.043	0.044	0.101	0.106	0.056	0.060	0.64	0.68	0.23		
<i>Buna S</i>											
25°	0.036	0.031	0.093	0.094	0.048	0.048	0.87	0.92		8.1	17.2
43°	0.039	0.036	0.093	0.096	0.050	0.053	0.69	0.72		6.6	9.8
<i>Perbunan</i>											
25°	0.028	0.026	0.079	0.084	0.035	0.038	1.28	1.38		13.5	48.0
43°	0.031	0.029	0.080	0.084	0.040	0.041	0.98	1.08		8.0	23.5
<i>Neoprene G</i>											
25°	0.029	0.026	0.075	0.075	0.036	0.036	0.83	0.83		8.8	18.1
43°	0.032	0.030	0.076	0.078	0.038	0.037	0.64	0.66		6.7	10.5
<i>Oppanol B 200</i>											
25°	0.035	0.034	0.107	0.110	0.052	0.049	0.69	0.69		1.25	3.6
43°	0.040	0.041	0.102	0.108	0.054	0.052	0.57	0.57		1.05	2.4
<i>Butadiene rubber</i>											
25°		0.034		0.094		0.045		0.97			
43°		0.039		0.097		0.049		0.82			
<i>Methyl rubber</i>											
25°		0.033		0.114		0.045		0.87			
43°		0.035		0.117		0.050		0.72			
<i>Mipolam MP</i>											
25°		0.022		0.052				1.00			
43°		0.027		0.059				0.80			
<i>Thiokol B</i>											
25°		0.011						0.30			
43°		0.013						0.28			

order to ascertain whether the quantity of permeating gas is proportional to the pressure and in inverse ratio to the thickness, as stated in the literature.⁸ This proved to be correct, in measurements made between 160 and 1700 mm Hg pressure with membranes of 0.20 to 1.01 mm thickness.

TABLE III. The specific gravity and the cubic expansion coefficient λ of various elastomers.

	Sp. gr. at 15°C	λ between 15° and 40°C
Natural rubber (vulcanized)	0.9689	7.0×10^{-4}
Buna S (vulcanized)	0.9974	6.7
Perbunan (vulcanized)	1.0316	6.2
Neoprene G (vulcanized)	1.2470	6.2
Oppanol B 200	0.9306	6.2

⁸ J. D. Edwards and S. F. Pickering, Bur. Stand. Sci. Paper 387 (1920).

The results of the main work are recorded in Tables I and II.

The thickness of the membranes, which varied from approximately 0.2 to 1.1 mm, was always so chosen as to produce a measurable time-lag. Of the gases used, H₂, O₂, N₂, SO₂, and NH₃ were obtained from cylinders, the gas CO₂ being taken from a Kipp apparatus. The methane was prepared by allowing water to drip into the Grignard compound of CH₃I with Mg.

To determine the solubility by the direct method it is necessary to know the specific gravity and the cubic expansion coefficient of the rubbers used. The specific gravity was therefore determined at a number of different temperatures by the buoyancy method, *vis.*, by weighing on an analytical balance, the rubber, suspended

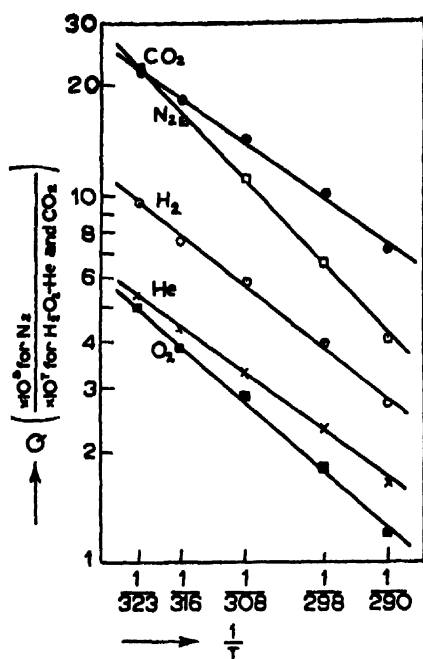


FIG. 4. Relation between the permeability in natural rubber and temperature.

on a hook, being dipped under water. The results are shown in Table III.

The specific gravity is accurate to 0.001 and λ to 0.2×10^{-4} . Within the measured range the coefficient of expansion is not a function of the temperature.

The computed solubilities contain the error of the permeability and diffusion determinations. To reduce this error to a minimum, the solubility was not calculated direct from the quotient of the permeability and diffusivity, but the latter quantities were first plotted logarithmically against $1/T$, as in Figs. 4 and 5. From this the solubility was calculated graphically. This was a precautionary measure against two accidental errors producing a considerable error in the solubility. As these solubilities are accurate to 10 percent, the results obtained by the direct method are the more reliable.

The permeabilities for H_2 and CO_2 at $25^\circ C$ agree within the experimental error with the permeabilities determined by other means in the presence of air.⁶ It follows from this that it is immaterial whether there is a vacuum or some alien gas on the other side of the membrane, as was the case in the former experimental tests. Thus per-

meability to a given gas is not affected by the presence of an alien gas. There is, likewise, agreement between solubilities and diffusivities, which were determined in two ways in this work.

For practical purposes the permeabilities at $25^\circ C$ of the various elastomers are calculated with reference to natural rubber in Table IV.

According to our results, the differences in permeability to a given gas are caused mainly by differences in rate of diffusion and only in a very minor degree by differences in solubility.

In Table V the permeabilities at $25^\circ C$ have been referred to H_2 . The differences in the permeability of a given elastomer to various gases are caused not only by great differences in rate of diffusion, but also to a considerable extent by differences in solubility. For example, the permeability to CO_2 is as a rule far greater than to H_2 owing to the far greater solubility of CO_2 , which amply compensates for its far slower rate of diffusion.

6. CALCULATION OF THE ACTIVATION ENERGY OF THE PERMEATION, THE ACTIVATION ENERGY OF THE DIFFUSION AND THE HEAT OF SOLUTION

The thermal dependence of the permeability and the diffusivity takes an exponential course, as it appears that there is a linear relation between $\log Q$ or $\log D$ and $1/T$. Some of the results plotted in this way are shown in Figs. 4 to 6, from which the applicability of the fol-

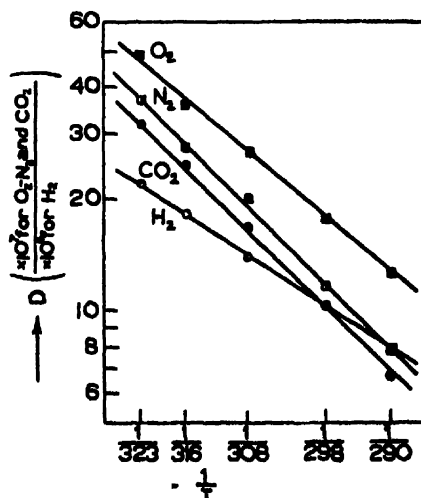


FIG. 5. Relation between the diffusivity in natural rubber and temperature.

lowing formulas may be concluded:

$$Q = Q_0 e^{-W/RT} \quad (11)$$

and

$$D = D_0 e^{-E/RT} \quad (12)$$

where Q_0 and D_0 = constants, W = activation energy of the permeation, and E = activation energy of the diffusion.

The slope of the lines in Figs. 4 to 6 gives a measure of the energy. The constants and the energies can be calculated with the aid of formulas (13) and (14), which follow from formula (12).

$$E = \frac{4.573 \log D_2/D_1}{1/T_1 - 1/T_2}, \quad (13)$$

$$\log D_0 = \log D_1 + \frac{E}{4.573 T_1}. \quad (14)$$

Dividing (11) by (12), we get

$$\frac{Q}{D} = \frac{Q_0}{D_0} e^{(-W+E)/RT}, \quad (15)$$

or

$$h = h_0 e^{(-W+E)/RT}. \quad (16)$$

And it is indeed a fact that the solubilities found which, in the case of natural rubber, were

TABLE IV. The permeability of elastomers at 25°C with reference to natural rubber.

	H ₂	O ₂	N ₂	CO ₂	CH ₄	He
Natural rubber	100	100	100	100	100	100
Butadiene rubber	82	80	74	103	—	—
Buna S	78	72	73	92	73	76
Methyl rubber	33	9	5.5	5.6	2.7	48
Perbunan	29	18	14	23	11	38
Neoprene G	26	17	14	19	11	—
Oppanol B 200	12	5	3.3	3.7	2.5	24
Mipolam MP	11	4	3	3.9	—	—
Thiokol B	3	1.2	—	2.4	—	—

TABLE V. The permeability of elastomers at 25°C with reference to H₂.

	H ₂	O ₂	N ₂	CO ₂	CH ₄	He
Natural rubber	100	46	17	260	56	59
Butadiene rubber	100	45	15	330	—	—
Buna S	100	43	16	310	52	57
Methyl rubber	100	12	2.8	44	4.6	85
Perbunan	100	28	7.7	200	21	76
Neoprene G	100	29	8.6	190	24	—
Oppanol B 200	100	18	4.5	77	11	114
Mipolam MP	100	16	4.5	91	—	—
Thiokol B	100	18	—	200	—	—

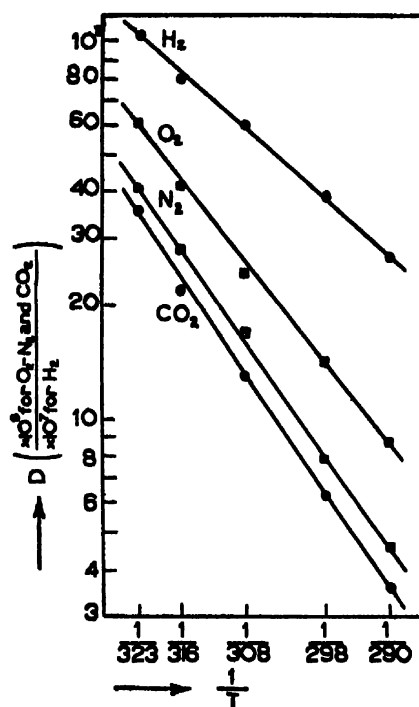


FIG. 6 Relation between the diffusivity in methyl rubber and temperature.

determined in complete thermal dependence, prove to take an exponential course with the temperature.

The following formulas hold for the calculation of the heat of solution:

$$\frac{d \ln C_B/C_1}{dT} = \frac{H}{RT^2}, \quad (17)$$

where H = heat of solution, C_B = concentration of the gas in the rubber, and C_A = concentration of the gas in the gas phase,

$$\frac{C_B}{C_A} = K_0 e^{-H/RT}. \quad (18)$$

Therefore, to calculate this heat of solution, the proportion of the concentration of dissolved gas to the concentration of the gas in the gas phase must be known, i.e., the number of cu. cm of gas of the prevailing temperature and pressure which is dissolved per cu. cm of solvent. It is quite simple to calculate this by multiplying our solubility figures, which are expressed in cu. cm of gas of 0°C and 76 cm by $T/273$.

The calculation then becomes as follows:

$$\ln h \frac{T}{273} = -\frac{H}{RT} + \ln K_0, \quad (19)$$

$$\frac{d \ln h}{d(1/T)} = -\frac{H}{R} + T = -\frac{H'}{R}, \quad (20)$$

$$H = RT + H'. \quad (21)$$

Thus, to find the required heat of solution, one can multiply the solubility h by $T/273$ and plot the logarithm of this amount against $1/T$. Alternatively, the logarithm of the solubility h can simply be plotted against $1/T$ and RT cal. added to the "heat of solution" H' thus obtained.

RT cal. has likewise to be added to $W - E$ to calculate the heat of solution from the temperature function of the permeability and diffusivity.

The heat of solution obtained in this way represents the supply of heat necessary to dissolve 1 mole of gas in rubber saturated with gas.

Table VI shows the activation energy of the permeation W for the various processes of permeation with the conjoined constant Q_0 . Measurements made by Barrer⁸ and Sager⁹ show satisfactory agreement with these figures. It will be evident that both quantities have little intrinsic value. But the energy W was used for the calculation of the heat of solution, and it also permits of the calculation of the permeability Q for any desired temperature. Caution is, however, necessary when calculating the permeability in this way at temperatures well outside those of our measured range, as the energy W may be a function of the temperature; indeed, with CH_4 , this already proved to be the case within that range, for which reason the activation energy W of CH_4 was not calculated. At lower temperatures the increase in permeability with the temperature was greater than at higher temperatures. With CO_2 as well, a similar tendency to depart from the purely exponential course was observed within the measured range, but with this gas the discrepancies were usually too small to affect the calculation of the activation energy W .

Table VI also gives the activation energy of the diffusion and the heat of solution of gases in

TABLE VI. The activation energies W and E with the conjoined constants Q_0 and D_0 and the heat of solution H in cal./mole.

		Q_0	W	D_0	E	H
Natural rubber I (from Q and D determination)	H_2	0.046	6900	0.23	5900	1600
	O_2	0.080	6600	0.57	7500	800
	N_2	0.47	9300	2.9	8700	1200
	CO_2	0.036	6200	3.6	8900	-2100
	He	0.014	6500	—	—	—
Natural rubber II (from Q and h determination)	H_2	—	—	0.31	6100	1400
	O_2	—	—	0.66	7600	700
	N_2	—	—	2.8	8600	1300
	CO_2	—	—	3.3	8800	-2000
	CH_4	—	—	—	—	-1300
Buna S	H_2	0.031	6800	0.056	5100	2300
	O_2	0.030	7300	0.23	7100	800
	N_2	0.12	8700	0.55	7800	1500
	CO_2	0.014	5700	0.90	8100	-1800
	He	0.013	6600	—	—	—
Perbunan	H_2	0.099	8100	0.96	7300	1400
	O_2	0.25	9400	2.6	9300	700
	N_2	1.1	11000	7.2	10200	1400
	CO_2	0.29	8300	12	10700	-1800
	He	0.012	7000	—	—	—
Neoprene G	H_2	0.090	8100	0.28	6600	2100
	O_2	0.54	9900	3.1	9400	1100
	N_2	0.56	10600	9.3	10300	900
	CO_2	0.33	8500	20	10800	-1700
	He	0.035	7900	—	—	—
Oppanol B 200	H_2	0.47	9500	0.55	7600	2500
	O_2	0.66	10700	7.3	10800	500
	N_2	5.5	12800	18	11700	1700
	CO_2	0.98	10100	22	11700	-1000
	He	0.035	7900	—	—	—
Butadiene rubber	H_2	0.022	6600	0.053	5100	2100
	O_2	0.023	7100	0.15	6800	900
	N_2	0.053	8200	0.22	7200	1600
	CO_2	0.0069	5200	0.24	7300	-1500
	He	0.0078	6600	—	—	—
Methyl rubber	H_2	0.10	8000	1.3	7300	1100
	O_2	3.3	11300	20	11100	800
	N_2	22	13300	105	12400	1500
	CO_2	9.7	11200	160	12800	-1000
	He	0.0078	6600	—	—	—
Mipolam MP	H_2	0.15	8900	0.20	6800	2700
	O_2	3.2	11800	4.3	10200	2200
	CO_2	0.68	9800	27	12000	-1600
	He	0.0078	6600	—	—	—
	He	0.0078	6600	—	—	—
Thiokol B	H_2	0.71	10600	2.0	8500	2700
	O_2	18	13500	—	—	—
	CO_2	6.9	11500	67	12100	0
	He	0.0078	6600	—	—	—
	He	0.0078	6600	—	—	—

the elastomers under test. In the case of natural rubber, moreover, the heat of solution is calculated from the direct determination of the solubility. The activation energy of the diffusion followed from that heat of solution together with the activation energy of the permeation. The figures resulting from these two forms of calculation agree.

The determination of the activation energies of the permeation and of the diffusion is accurate to 200 cal. The reason why this is so much more accurate than measurement of the permeability and diffusivity alone is that the calculation of these energies does not necessitate the use of various particulars which themselves contain errors, like the thickness and area of the membrane and volume of the apparatus, since all

⁹ F. P. Sager, J. Research Nat. Bur. Stand. 25, 309 (1940).

that is needed to calculate W is to plot the logarithm of the pressure increase per unit of time (after conversion to 0° and 76 cm) against $1/T$. It likewise suffices for the calculation of E to plot the logarithm of the time-lag against $1/T$. In unfavorable cases, i.e., when the errors in W and E reinforce each other, the determination of the heat of solution is faulty to 400 cal. The accuracy of Q_0 and D_0 is entirely governed by that of W and E . An error of 200 cal. in W or E makes Q_0 or D_0 faulty by approximately 30 per cent.

7. DISCUSSION OF THE SOLUBILITY

With the data we have collected, we shall now consider more closely the two factors which govern permeability, *viz.*, solubility and rate of diffusion, in an endeavor to find the reason for the magnitude of the permeability of rubber to gases. It will become apparent that the behavior of gases in rubber is entirely analogous to that of gases in organic liquids. So, rubber may be regarded as an organic liquid of very high molecular weight.

One of the striking facts which emerges from the solubility measurements is that the solubility of H_2 and N_2 in rubber-like substances increases as the temperature increases, from which it follows that heat must be supplied for the process of dissolving. This is contrary to Barrer's⁸ experience upon determining the solubility of H_2 and N_2 in some elastomers, when the solubility decreased as the temperature rose. Apart from the fact that we established the dependence of the solubility upon the temperature in two unallied ways, our finding is the more probable because the solubility of these gases in organic liquids almost invariably increases together with the temperature.¹⁰ More soluble gases, like CO_2 for example, always exhibit a decrease in solubility, on the contrary, in organic liquids as the temperature rises.

The explanation we suggest for this behavior of gases is the following:¹¹

The heat of solution H consists of a thermal value Q , depending upon the chemical or physical

action between gas and liquid molecules, and of a work A brought about by the passing of 1 mole of gas into the solution, hence

$$H = A + Q, \quad (22)$$

whereby the heat absorbed during solution and the work then exerted by the gas become positive factors in the calculation. Furthermore

$$A = RT \ln P/p, \quad (23)$$

A being regarded as the work produced upon the conversion of 1 mole of gas from outside pressure P to the osmotic gas pressure p in the liquid. It should be noted that p/P is the solubility of the gas expressed as cu. cm gas of the prevailing temperature dissolved per cu. cm rubber.

If the concentration of the gas in the liquid is smaller than outside, A is positive. If Q is neglected, then $A = H$, with the result that the heat of solution H , which has to be supplied, is, likewise, positive, and solubility rises as the temperature rises.

If, on the other hand, the concentration of the gas in the liquid is greater than outside, A will be negative. Neglecting Q , it follows from this that the heat of solution H is negative and that the solubility will decrease as the temperature rises.

Reference to Table VI will show that the heat of solution of a gas in elastomers is less positive or more negative in proportion to greater solubility. It has been found that the heat of solution of N_2 and H_2 in the various rubbers agrees fairly closely with the work of expansion as calculated from (23). The thermal value Q cannot be neglected in the case of the other gases. Taking CO_2 as an instance, its solubility is of a kind that the concentration of the gas inside and outside the rubber is about equal; hence the work A is very small, yet an appreciable negative value is found for the heat of solution.

The degree of solubility is governed by the nature of the gas and of the elastomer.

a. The Nature of the Gas

The fact is known that a gas with a high critical temperature or boiling point dissolves more readily in an organic liquid than a gas of

¹⁰ Cf J. Horiuti, Sci. Pap., Inst. Phys. Chem. Research, Tokyo 17, 125, 256 (1931).

¹¹ G. Tammann, Zeits. f. anorg. Chemie 158, 17 (1926); 194, 159 (1930).

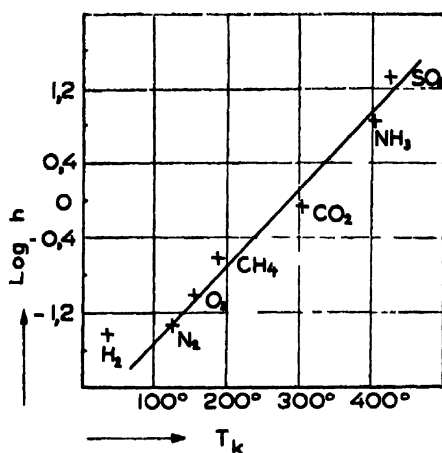


FIG. 7. Relationship between the critical temperature and solubility of gases in natural rubber.

lower critical temperature or boiling point.¹² The position of the critical temperature is a measure of the tendency of the gas to condense, with which the solubility stands in relation.

With rubbers, too, the solubility of the gas increases with the critical temperature.¹³ As may be seen from Fig. 7, a relation exists between the solubility h and the critical temperature of the gas T_k :

$$\log h = A + BT_k, \quad (24)$$

where A and B are constants.

H₂ is somewhat more soluble in rubber than would seem to follow from (24).

Körösy¹⁴ found the same relation for the solubility of gas in organic liquids, with H₂ showing a similar deviation.

b. The Nature of the Elastomer

The presence of polar groups reduces the solubility of a non-polar gas in an organic liquid,¹⁵ for two electrically polar molecules will have greater mutual attraction than two which are not polar. In a solution of non-polar molecules in a polar solvent this greater mutual attraction will more or less expel the non-polar molecules. Contrarily, polar gas molecules will dissolve

more readily in a liquid with polar molecules than in a liquid without polar molecules.

All this is equally applicable to rubber-like substances. Perbunan with strong polar groups (—CN groups) has a reduced solvent power with respect to non-polar gases—H₂, N₂ and O₂—and an increased solvent power upon the polar gases SO₂ and NH₃, while on the other hand the saturated hydrocarbon Oppanol has very low solvent power over the polar gases. Perbunan also has an increased solvent power with respect to CO₂, despite the fact that this gas has no dipole moment as have SO₂ and NH₃. Nevertheless, CO₂ is not non-polar, for it contains two polar groups, which neutralize each other exteriorly,¹⁶ one result of which is that, on the whole, CO₂ dissolves more satisfactorily in a substance with polar groups than in one without them, so in this case in Perbunan more than in natural rubber. The same thing is not found, however, with Neoprene (containing —Cl groups).

8. DISCUSSION OF THE ACTIVATION ENERGY OF DIFFUSION

The value of the diffusivity is governed by the value of the constant D_0 and the activation energy of the diffusion E . The activation energy E has to be supplied to separate the rubber chains—which are linked together by Van der Waals forces¹⁷—a sufficient distance to permit a gas molecule to be displaced or to make for itself a new place or hole. This energy is supplied by the kinetic energy of the gas molecule and particularly by the kinetic energy of the rubber molecules.

The degree of activation energy is governed by the nature of the gas and of the elastomer.

a. The Nature of the Gas

It might be expected that the larger the diameter of the gas molecule, the greater will the activation energy be, since larger openings between the rubber molecules will be needed for the diffusion of the larger gas molecules; consequently, these rubber molecules will need to be separated over a greater length, for which more energy is required.

¹² J. H. Hildebrand, *Solubility of Non-Electrolytes* (Reinhold Publishing Corporation, New York, 1936), second edition, p. 32.

¹³ Cf. C. S. Venable, and T. Fuwa, *Ind. Eng. Chem.* 14, 139 (1922).

¹⁴ F. Körösy, *Trans. Faraday Soc.* 33, 416 (1937).

¹⁵ J. H. Hildebrand, reference 12, p. 133.

¹⁶ G. Briegleb, *Zwischenmolekulare Kräfte und Molekülstruktur* (F. Enke Verlag, Stuttgart, 1937), pp. 50 and 61.

¹⁷ R. H. Ewell, *J. App. Phys.* 9, 252 (1938).

Figure 8 shows the activation energy in relation to the diameter d_T of the gas molecules used,¹⁸ which was determined from the internal friction. It appears that the activation energy of the diffusion approximately increases in proportion to the size of the gas molecules.

b. The Nature of the Elastomers

It is to be expected that the stronger the secondary valencies by which the rubber molecules are interlinked, the greater will be the activation energy of the diffusion E . These secondary valencies also govern the extent of the heat of evaporation and the cohesive energy of the molecules.

The introduction of methyl groups into rubber causes the cohesive energy to increase, as may be seen in Table VII, where each successive elastomer has one methyl group per structural unit more than the previous elastomer. It is evident that the activation energy and the cohesive energy are interrelated. This cohesive energy applies per unit of structure and is calculated by Dunkel's method.¹⁹ We may conclude that approximately $1\frac{1}{2}$ to $2\frac{1}{2}$ units of structure have to be separated to allow the diffusion of one gas molecule.

Oppanol is another instance of the considerable importance of the methyl groups in the molecule, for in it, as in methyl rubber, the activation energies for the diffusion were found to be very high. It is not possible in this case to calculate the cohesive energy following Dunkel, because the increment for this energy of a quaternary C atom is not known, but, as compared with natural rubber, Oppanol may be expected to possess considerable cohesive energy per linear unit of the chain molecule, as the Oppanol molecules contain a great many methyl groups.

TABLE VII. Comparison between the cohesive energy and the activation energy E (in cal./mole).

	Cohesive energy	E_{H_2}	E_{O_2}	E_{N_2}	E_{CO_2}
Butadiene rubber	3960	5100	6800	7200	7300
Natural rubber	4370	5900	7500	8700	8900
Methyl rubber	4780	7500	11100	12400	12800

¹⁸ H. A. Stuart, *Molekulstruktur* (Verlagsbuchhandlung Julius Springer, Berlin, 1934), p. 36.

¹⁹ M. Dunkel, *Zeits. f. physik. Chemie* A138, 42 (1928).

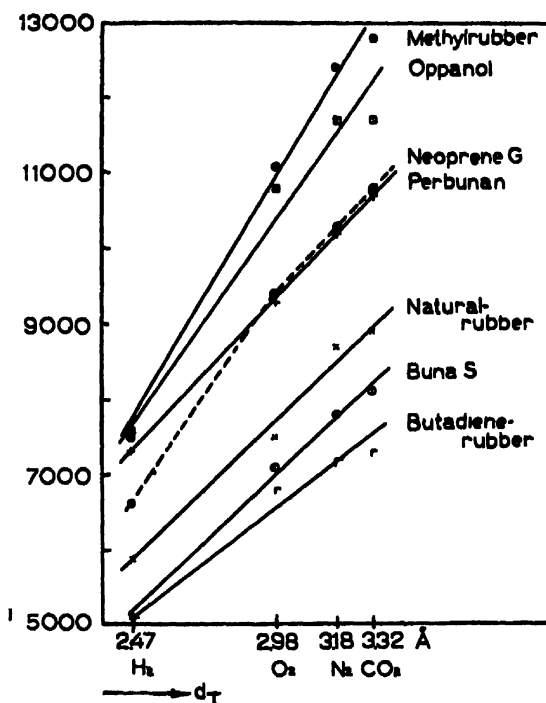


FIG. 8. The activation energies of the diffusion in relation to the diameter of the gas molecules.

The introduction of polar groups greatly increases the cohesive energy of molecules. Accordingly, a high activation energy of the diffusion is found in Neoprene and Perbunan as compared with butadiene rubber.

To sum up, it may be said that for the investigated materials a considerable activation energy of the diffusion, and thereby a low rate of diffusion and low permeability, may be brought about by the presence in the elastomer of (a) polar groups, e.g., Perbunan, Neoprene, Mipolam, and Thiokol; (b) methyl groups, e.g., methyl rubber and Oppanol.

9. DISCUSSION OF THE CONSTANT D_0

Several attempts have been made in the literature²⁰ to relate the temperature function of the diffusivity to other quantities. Most of the formulas describing this relation are of the following type:

$$D = \frac{1}{2} \nu d^2 e^{-E/RT}, \quad (25)$$

where ν is a frequency and d a length.

²⁰ R. S. Bradley, *J. Chem. Soc.* 1910 (1934); *Trans. Faraday Soc.* 33, 1185 (1937). R. M. Barrer, *Proc. Phys. Soc. London* 52, 58 (1940).

The factor $\frac{1}{2}\nu d^2$ then corresponds to the constant D_0 as determined by us. If the frequency ν is conceived as the number of vibrations executed by the gas molecule in the rubber and the length d as the free length of path of this gas molecule in the rubber, it would follow that D_0 must be roughly constant in the diffusion of a given gas in various rubbers. Actually there is great disparity between the D_0 's. It was found, for example, that for N_2 $D_0=0.22$ to 105, depending very much upon the activation energy E .

This makes it plain that a formula of this type does not hold good for rubber.

There is another formula, suggested by Barrer,²¹ which might meet the case, viz.,

$$D = \frac{1}{2}\nu d^2 \frac{(E/RT)^{f-1}}{(f-1)!} e^{-E/RT}, \quad (26)$$

f = number of degrees of freedom.

This formula only applies if $(E/RT) \ll f-1$.

If this formula is applied to our results, then E has been wrongly calculated, since the fact that D_0 may depend upon the temperature has not been taken into account, for

$$\frac{d \ln D}{d(1/T)} = -\frac{E}{R} + (f-1)T = \frac{-E(\text{determined})}{R}. \quad (27)$$

Therefore, $(f-1)RT$ has to be added to the E already determined and this new E must be filled in in (26). Barrer worked out the results of his diffusion measurements of gas in rubber with this formula and came to the conclusion that, to obtain a proper order of magnitude for D_0 , the number of degrees of freedom must amount to 12 to 14.

We worked out some of our results with this formula, assuming that $\nu = 5 \times 10^{12}$ and that $d = 8\text{\AA}$. It makes no fundamental difference if some other amount is taken for these quantities.

TABLE VIII. Calculation of the number of degrees of freedom f in accordance with formula (26).

	D_0	E	Calculated
H_2 —Buna S	0.056	5100	$f=4$ $D_0=0.051$
H_2 —Natural rubber	0.23	5900	6 0.21
H_2 —Perbunan	0.96	7300	8 0.93
N_2 —Oppanol	18	11700	10 22
N_2 —Methyl rubber	105	12400	13 96

²¹ R. M. Barrer, Trans. Faraday Soc. 35, 644 (1939).

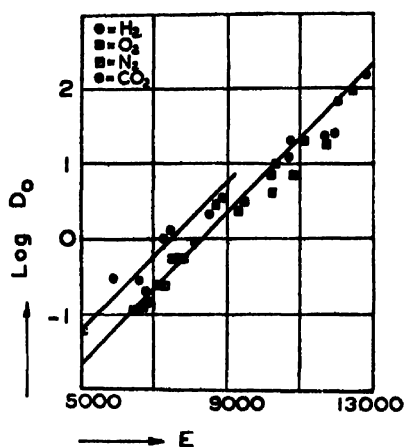


FIG. 9. Relation between $\log D_0$ and E .

It appears from Table VIII that the number of degrees of freedom may vary from 4 to 13, according to the activation energy, which does not clarify the problem at all. It is difficult to avoid the impression that f here is little more than a factor by which D_0 may be brought into the right order of magnitude.

So, as this course does not provide satisfactory information on the magnitude of D_0 , we tried to draw up an experimental formula based on the great dependence of D_0 upon E . For this purpose all the D_0 values found were logarithmically plotted against E (see Fig. 9), with the result that a practically linear relation is found to exist between $\log D_0$ and E . Barrer²² also found this relation.

The points in Fig. 9, fall fairly clearly into two groups. The first group, with O_2 , N_2 , and CO_2 , follows by approximation a line which conforms to the equation

$$\log D_0 = E/2000 - 4.2$$

or

$$D_0 = 6.3 \times 10^{-5} e^{E/870}. \quad (28)$$

The second group with H_2 conforms by approximation to

$$\log D_0 = E/2000 - 3.75$$

or

$$D_0 = 1.8 \times 10^{-4} e^{E/870}. \quad (29)$$

It also follows from these formulas that H_2 diffuses on an average about 2 to 3 times more

²² R. M. Barrer, Trans. Faraday Soc. 38, 322 (1942).

quickly than O₂, N₂, or CO₂, apparently on account of the greater average velocity of H₂ molecules. The kind of rubber also seems to exert a specific influence, as gas diffuses in natural rubber approximately twice as quickly and, in Oppanol and Mipolam, about twice as slowly as is to be inferred from the formulas. The combination of the formulas (28) and (29) with $D = D_0 e^{-E/RT}$ leads for N₂, O₂, and CO₂ to

$$D = 6.3 \times 10^{-8} \exp \left[-E \left(\frac{1}{RT} - \frac{1}{870} \right) \right], \quad (30)$$

and for H₂ to

$$D = 1.8 \times 10^{-4} \exp \left[-E \left(\frac{1}{RT} - \frac{1}{870} \right) \right]. \quad (31)$$

Hence these formulas apply by approximation to all species of rubber and the first one even to three gases. According to (30) and (31), the diffusivity D depends only, with a given gas, upon E and the temperature; thus any random rubber could be roughly classified merely by measuring the rate of diffusion at a certain temperature. Indeed, with only one exception, the higher activation energies E were always found to go along with the lower diffusivity D .

There would be some justification for concluding from this reasoning that the considerable differences in the e -power are responsible for the

great differences in the D_0 's found and it may therefore be asked whether the differences in D_0 have any real significance.

The most serious problem with which one is faced when drawing up a formula intended to meet all theoretical and experimental demands is that of accounting for the heat motion of the rubber molecules, and the failure of diffusion formulas which do satisfy other diffusion processes is probably caused by this inadequate allowance for the heat motion. This view accords with the results of an investigation carried out by Kuhn²² into the rate of diffusion of organic dyes in swollen and de-swollen rubber, where, likewise, the heat motion of the rubber molecules proved to be of decisive importance.

ACKNOWLEDGMENT

In conclusion, I wish to express my thanks to the Rubber-Foundation for permission to publish the results of the investigation; to Professor Dr. A. van Rossem for his interest and advice; and to Professor Dr. C. Zwikker for his collaboration in the preparation of the thesis which was published in Delft in 1943 and of which this paper reproduces a part.

²² W. Kuhn, H. Suhr, and K. Rijffel, *Helv. phys. acta* **14**, 497 (1941).

Here and There

New Appointments

Donald Kirkham has accepted a position as associate research professor in the Department of Agronomy and Physics at Iowa State College, where he will be in charge of the soil physics work.

William H. Myers has been appointed chief engineer of the receiver division, Farnsworth Television and Radio Corporation.

Carl J. Christensen of the Bell Telephone Laboratories technical staff has been appointed dean of the School of Mineral Industries at the University of Utah.

R. O. Curry is now audio and acoustical engineer for the Farnsworth Television and Radio Corporation. He was previously engaged in audio research for the Capehart Division of the company.

Willard R. Beye, formerly senior engineer at Transcontinental and Western Airlines, Kansas City, Kansas, has joined the staff of the Midwest Research Institute. He will work in the design and development section. Y. W. Luke, computer and mathematician, has also become associated with the Institute.

Howard P. Wile is the first appointee to the new post of administrator of research for contract projects at the Polytechnic Institute of Brooklyn.

Frank H. Spedding has been appointed director of the newly-established Atomic Research Institute at The Iowa State College, Ames, Iowa.

New Section for Edison Laboratory

The Edison Laboratory, central research organization of Thomas A. Edison, Incorporated, West Orange, New Jersey, has announced the formation of a new laboratory section for fundamental research in the branches of science which are the bases of the company's business in dictating machines, batteries, electrical instruments, and other electrical devices. The technical staff of the new group includes Dr. Robert S. Weisz, inorganic chemist, Dr. Richard W. Iskraut, physicist, S. S. Jaffe, chemist, and Dr. Robert L. Tichenor, physical chemist.

Radio Technical Planning Officers

The following men have been elected to serve as officers of the Radio Technical Planning Board for the year beginning October 1: chairman, HAROLD PRATT of Mackay Radio and Telegraph Company; vice chairman, J. L. MIDDLEBROOKS of the National Association of Broadcasters, Inc.; secretary, GEORGE W. BAILEY, executive secretary of the Institute of Radio Engineers, Inc.; treasurer, WILL BALTIMORE of the Television Broadcasters Association.

Bausch and Lomb Scholarships Awarded

Competition among fifteen finalists for five Bausch and Lomb Science Scholarships at the University of Rochester, valued at \$1500 each, was so keen that the University has awarded a total of eleven scholarships, with an aggregate maximum value of more than \$16,000. The finalists were chosen from nearly 450 contestants among the highest ranking high school science students in the United States, and were brought to Rochester, New York, for final tests and interviews. This was the third annual nationwide competition for these scholarships. Eight boys and three girls were the winners.

Appeal for Better Prosthetic Devices

Dr. Paul E. Klopsteg, chairman of the Committee on Prosthetic Devices, has issued, through the facilities of the National Inventors Council, a challenge to American inventors and mechanical engineers to develop better artificial limbs. Stressing the urgent needs of 17,000 war amputees and an estimated annual minimum of 25,000 amputees, Dr. Klopsteg urged inventors and engineers to submit ideas and suggestions for better devices to John C. Green, Chief Engineer, National Inventors Council, Department of Commerce, Washington 25, D. C. He listed the following pressing problems as still requiring solution: improved joint mechanisms, lighter materials, new fabrication methods, fundamental study of the mechanics of human motion, improved fitting procedures.

New Officers of Illuminating Society

Gilbert K. Hardacre, manager of commercial sales of the Public Service Company of Northern Illinois, has been elected president of the Illuminating Engineering Society. Other new officers are: general secretary, Walter Sturrock; treasurer, Charles H. Goddard; vice president, Lee E. Taylor; directors, Allan E. Parker and P. M. Rutherford, Jr.

Meeting of Canadian Electron Microscopists

From June 12 to 14 representatives of the four Canadian electron microscope laboratories (University of Toronto, McGill University, National Research Council, and Shawinigan Chemicals, Ltd.) met with the research staff of Shawinigan Chemicals, Ltd., at Shawinigan Falls for presentation of papers concerning their own work and discussion of techniques and problems.

The Bulletin of Mathematical Biophysics

SEPTEMBER 1946

Some Considerations on the Structure of the Quadruped Extremities—N. RASHEVSKY
Tree Trunks and Branches as Optimum Mechanical Supports of the Crown: II. The Branches—MARTINUS H. M. ESSER
The Probabilistic Approach to the Effects of Radiations and Variability of Sensitivity.—I. OPATOWSKI
On the Relationship Between Response Time and Dosage of a Drug as a Function of its Mode of Entry—H. D. LANDAUER

The University of Chicago Press, Chicago, Illinois, Vol. 8, No. 3.

Joint Meeting of EMSA and ASKRED

On December 5, 6, and 7, 1946, a joint meeting of the Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction will occur at Mellon Institute and the University of Pittsburgh. All interested persons are welcome to attend. The program is as follows:

Thursday, December 5, 9:30 a.m. (Joint session)

1. Trends in x-ray diffraction work. DAVID HARKER, *General Electric Company*.
2. Trends in the use of the electron microscope in the study of fine structure. JAMES HILLIER, *Radio Corporation of America*.
3. The application of electron diffraction methods in the study of fine structure. ROBERT HEIDENREICH, *Bell Telephone Laboratories*.

2:00 p.m. (EMSA)

1. First-order theory of the three electrode electron gun with plane electrodes. S. G. ELLIS, *University of Toronto*.
2. Some observations on the shadow cast formvar replica technique. BEATRICE M. DRACON, S. G. ELLIS, AND W. G. CROSS, *University of Toronto*.
3. Electron microscopy of colloid systems. JOHN TURKEYVICH, *Princeton University*, AND JAMES HILLIER, *Radio Corporation of America*.
4. Electron microscope examination of the microphysical properties of the polymer cuprene. JOHN H. L. WATSON, *Shawinigan Chemicals Limited*.
5. An effect of electron bombardment upon carbon black. JOHN H. L. WATSON, *Shawinigan Chemicals Limited*.
6. The electron microscope examination of greases. S. G. ELLIS, *University of Toronto*.
7. Artifacts produced by poor resolution. I. H. WILLISFORD, *Goodyear Tire and Rubber Company*.

2:00 p.m. (ASKRED)

1. Diffraction patterns of some non-crystalline soap phases. H. NORDSIECK, F. B. ROSEVEAR, AND R. H. FERGUSON, *Procter and Gamble Company*.
2. Triple chain length structure in mixed saturated triglycerides. E. S. LUTTON, *Procter and Gamble Company*.
3. Preparation and diffraction studies of synthetic diacid diglycerides. B. F. DAUBERT AND S. S. SIDHU, *University of Pittsburgh*.
4. The crystal structure of the sugars. R. PEPINSKY, *Alabama Polytechnic Institute*.
5. Business meeting.

6:30 p.m.

Joint dinner and address by FRANCIS O. SCHMITT, *Massachusetts Institute of Technology*.

Friday, December 6, 9:00 a.m. (EMSA)

Symposium on Biological Problems

1. Remarks on the borderline of physics, chemistry, and biology. L. MARTON, *Stanford University*.
2. Applications of the phase microscope in electron microscopy. OSCAR W. RICHARDS, *American Optical Company*.
3. Preparation of tissue for high speed sectioning. MARY C. SCHUSTER AND CLIFFORD E. GRAY, *Interchemical Corporation*.
4. Electron microscope studies on the structure of chlorophyll "crystals," "crystalline alpha carotene," "crystalline beta carotene," and "crystalline vitamin A." EDITH A. ROBERTS, *Vassar College*.
5. Further electron microscope studies of the structure of starch grain. MILDRED D. SOUTHWICK, *Vassar College*.
6. The mounting of bacteria for electron microscope examination. JAMES HILLIER AND R. F. BAKER, *Radio Corporation of America*.
7. A study with the high-voltage electron microscope of the structure and germination of the endospores of *Bacillus mycoides*. GEORGE KNAYSI, R. F. BAKER, AND JAMES HILLIER, *Radio Corporation of America*.
8. Demonstration with the high-voltage electron microscope of nucleus-like bodies in young cells of *Bacillus mycoides* grown in nitrogen free

medium. GEORGE KNAYSI AND R. F. BAKER, *Radio Corporation of America*.

9. Observations on actin and myosin extracted from muscle. M. A. JAKUBS AND C. E. HALL, *Massachusetts Institute of Technology*.

9:00 a.m. (ASKRED)

1. An application of Fourier transforms to crystal structure analysis. DOROTHY WRINCH, *Smith College*.
2. Fourier synthesis of lattice disorders. W. H. ZACHARIASEN, *University of Chicago*.
3. Isothermal order—disorder in a salt system. ALEXANDER GRENALL, *Union Oil Company of California*.
4. The indexing of powder diffraction data. F. W. MATTHEWS, *Canadian Industries Limited*.
5. Ambiguities in diffraction analysis. A. L. PATTERSON, *Bryn Mawr College*.
6. A study of heterogeneous reactions by the electron diffraction and electron microscope techniques. EARL A. GULBRANSEN, *Westinghouse Research Laboratories*.

2:00 p.m. (EMSA)

Symposium on Biological Problems

1. Size distribution of tobacco mosaic virus in the early stages of infection. GERALD OSTER, *The Rockefeller Institute for Medical Research*.
2. Studies on the Newcastle disease virus (California strain). A. R. TAYLOR, *Duke University Medical School*.
3. An electron microscope study of the structures of some virus protein crystals. R. W. G. WYCKOFF, *National Institute of Health*.
4. Immunochemical aspects of winkle treated virus. SAUL MALKIEL, *The Rockefeller Institute for Medical Research*.
5. Business meeting.

2:00 p.m. (ASKRED)

1. Studies of a montmorillonite cracking catalyst. ALEXANDER GRENALL, *Union Oil Company of California*.
2. An x-ray diffraction study of a new cobalt carbide. L. J. E. HOWAR AND W. C. PREBLE, *U. S. Bureau of Mines*.
3. An x-ray diffraction study of chrome steel slags. G. P. CHATTERJI AND S. S. SIDHU, *University of Pittsburgh*.
4. Determination of silica in industrial dusts by means of the Geiger counter x-ray spectrometer. HAROLD P. KLUG, LEROY E. ALEXANDER, AND ELIZABETH BOLLMAN, *Mellon Institute*.
5. An x-ray diffraction study of the silver-magnesium alloy system. HAROLD R. LITNER AND S. S. SIDHU, *University of Pittsburgh*.
6. A long spacing x-ray diffraction camera. ORVIL BOLDUAN, *Massachusetts Institute of Technology*.

8:30 p.m. (Joint session)

Report on London Meeting

DAVID HARKER, *General Electric Research Laboratory*
I. FANKUCHEN, *Polytechnic Institute of Brooklyn*

Saturday, December 7, 9:00 a.m. (Joint session)

1. Specimen preparation for electron micrographs and diffraction patterns. LELAND L. ANTEN, *University of Texas*.
2. Alterations and additional equipment for an R.C.A. console electron microscope. HENRY C. FROULE, *Armour Research Foundation*.
3. Spherical aberration of compound magnetic lenses. L. MARTON AND K. BOL, *Stanford University*.
4. Further studies on the magnetic electron microscope objective. JAMES HILLIER AND E. G. RAMBERG, *Radio Corporation of America*.
5. The bias focused electron gun. JOHN H. REISNER, *Radio Corporation of America*.
6. Photographic plates for use in electron microscopy and electron diffraction. ROBERT G. PICARD AND W. F. SWANN, *Radio Corporation of America*.
7. Calculated magnitude of Fresnel diffraction in shadow casting. C. E. HALL, *Massachusetts Institute of Technology*.
8. A review of some recent advances in electron microscopy microtomy. E. FULLAM, *General Electric Research Laboratories*.
9. Round table discussion of cross sectioning for the electron microscope. Invited speakers and discussion from the floor.

Ohmite Laboratory Dedicated

The new Ohmite Laboratory was dedicated in June in an all-day ceremony at the Armour Research Foundation of Illinois Institute of Technology to serve industry of the Middle West with precision electrical measurements established on the basis of and approved by the Bureau of Standards.

Oak Ridge Groups Merge

The Association of Oak Ridge Scientists and the Association of Oak Ridge Engineers and Scientists have been dissolved in favor of a single organization, The Association of Oak Ridge Engineers and Scientists. The amalgamation became effective at a general meeting of the membership of both organizations held in June. For the present, correspondence to the new organization or any officer or member thereof may be addressed to either P. O. Box 382 or P. O. Box 1251. All debts, assets, and activities of the two older organizations have been transferred to the new group and its officers.

New Physics Division at Cornell

One of the nation's first schools of engineering physics with a five-year curriculum was opened at Cornell University in September, "to train students in a combination of engineering principles, applied physics and mathematics for careers in research and development." According to Dean S. C. Hollister of the College of Engineering through which the new department is being administered, "The Department of Engineering Physics will bridge the area lying between the training in basic science and in fundamental engineering." He pointed out that the traditional curriculum in engineering falls short on the side of basic science, and the physics curriculum falls short on engineering principles.

The degree of Bachelor of Engineering Physics will be awarded upon completion of the five-year curriculum, and graduate work will be offered through the Graduate School of the University. Dr. Lloyd P. Smith is serving as director of the new department in addition to his post of chairman of the Department of Physics.

Training in Electron Microscopy at Toronto

Professor E. F. Burton of the McLennan Laboratory Department of Physics, University of Toronto, has announced that a graduate course will be given next year for the purpose of training graduate students in all phases of electron microscopy theory and practice. The course next year will be limited to six students who must have had advanced training in mathematics and physics. The course will be accepted as work toward the degree of M.A., and candidates must be acceptable to the School of Graduate Studies, University of Toronto.

New Booklets

Split Ballbearing Corporation, Lebanon, New Hampshire, has issued Catalog No. 84, giving full specifications on sizes, load ratings, etc., on a complete line of divisible race, ball, roller, and thrust bearings. 30 pages. Free on request.

United Electric Controls Company, 69 A Street, Boston 27, Massachusetts, has published a new catalog giving information and engineering details on all their thermostats and pressure switches. 30 pages. Free on request.

H-B Instrument Company, 2518 North Broad Street, Philadelphia 32, Pennsylvania, announces its Catalog No. 14 listing temperature controls, relays, and thermometers. 16 pages. Free on request.

The "Milvay Notebook" for Spring, 1946, published by Chicago Apparatus Company, Chicago 22, Illinois, has the following table of contents:

Titrimetric Procedures.
Microscope Camera Attachment.
Spectrophotometer.
Synthetic and More Synthetics.
Glas Col Flask Heaters.

34 pages, numerous illustrations.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, recently issued a leaflet describing the L123 precision spectrometer with two microscopes, a camera attachment, polarizing attachment, Babinet and Babinet Soleil compensators, Abbe-Lamont auto-collimating eyepiece, filar eyepiece micrometer, and various other items. 8 pages. Free on request.

Eastman Kodak Company, Rochester 4, New York, has announced a third revised edition of the *Kodak Data Book on Formulas and Processing*. Containing a comprehensive list of Kodak formulas together with a discussion of the principles and procedures for processing films, plates, and papers, the new edition incorporates a number of changes to bring it up to date.

Bulletin 155-64 of The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, describes the new Gaertner large toolmaker microscope. 8 pages. Free if request is made on company letterhead.

General Electric Company's *Special Products Digest* for June 1946 features an article on the G-E 100-million-volt betatron and its use in atomic research. 4 pages.

The July-August issue of *Westinghouse Newsfront* contains a description of the newest methods of cleaning and drying indoor air, and a page of pictures of different uses of the servo-mechanism. 4 pages.

Journal of Applied Physics

Volume 17, Number 12

December, 1946

Report of the Electron Microscope Society of America's Committee on Resolution

I. INTRODUCTION

THE opinions expressed in this report are those of the committee members and are based upon the personal experiences and experiments of those members, the published literature on the subject, and the answers received to a questionnaire sent out to all members of the Electron Microscope Society. There was not sufficient time available for the Committee, as such, to carry out experimental tests of the ideas and methods reported here. However, such experimental tests should be carried out, either by this or a subsequent committee or by interested members of the Society.

A suggestion was received from some members of the Society that the Committee examine various methods of measuring "resolving power" and recommend to the Society that one, the best of these methods in the opinion of the Committee, be adopted as *the* method of the Society. Consideration was given this suggestion but the Committee concluded that, for reasons to be explained later, such a suggestion could not be carried out practically. Consequently, the Committee does not, at this time, especially recommend any specific method of measurement.

It was decided that the interests of the Society could best be served by a critical review of the entire subject of resolving power and resolution, and of the various known or suggested methods for quantitative or qualitative evaluation of this factor. This report, therefore, contains the efforts of the Committee in such a direction.

II. DISCUSSION OF THE TERMINOLOGY USED

To obtain increased resolution is the fundamental reason for using an electron microscope. Such a statement should be obvious, yet careful consideration of it indicates the importance of any discussion of resolution. For much of the work with the microscope it is sufficient to know merely that the resolving power is about 2 orders of magnitude better than that of light microscopes. For example, if the thirty replies to the questionnaire sent to the members of the Society by this Committee give a representative cross section, only 30 percent of the micrographs taken with the instrument are examined at a total magnification of 50,000 times or greater. Now, unless this range of magnification is employed, the limiting factor in the detail to be seen is the resolving power of 0.2 mm for the human eye, and about 40A for the microscope.

Therefore, it should be realized that in about 70 percent of the work with present-day microscopes the question of resolution may be neglected as a limiting factor. However, the feeling was brought out in several replies that a greater percentage of high magnification work would be done if the resolving power of the microscope warranted it.

For the 30 percent of the micrographs which were examined at magnifications above 50,000X, however, the resolution does become important. In this connection, it was of interest to the Committee, as it will be to the Society, to obtain information on what members had determined

or believed to be their maximum resolution. The results of these replies are given in Fig. 1.

It is difficult to believe that the wide variation in this value is caused by actual variabilities in the instruments. Much of the differences must be caused by the lack of exact nomenclature associated with the subject. This may perhaps be traced to the fact that definitions commonly encountered are usually purely descriptive and invoke the concept of abstract mathematical points whose relation to physical particles is not always indicated. Furthermore, it is the opinion of the Committee that there are several aspects of resolutions or resolving power which are significantly different and should not be directly compared.

As examples of the descriptive types of definition, we may use those given by Webster's International Dictionary as follows:

"Optical resolving power—The power of an optical system using light of one color to form images of two point sources which are close together, so that the images may be recognized as distinct."

"Resolving power—The ability of an optical instrument to form distinct images of objects separated by small angular distances."

"Resolution—The act or property of rendering visible the separate parts of an object or of distinguishing between different sources of light."

Resolving power has also frequently been defined in the literature as the minimum distance between two points of the object whose images can still be resolved as separate.

It is the first recommendation of the Committee that the two terms, resolving power and resolution, be carefully distinguished in the following manner. As implied in the term "power" as well as in the Webster definition just given, *resolving power* should always be considered as an *ability* of the optical system while the term *resolution* should connote only the *realization* of that ability in a specific image. Resolving power should be the maximum resolution which *can be obtained* with the system. The actual resolution present in any specific image or picture may well be somewhat poorer. Thus, to speak of a series of published micrographs as having been obtained with an instrument whose resolving power is 20A is misleading. The actual resolution in the pictures shown or even possible

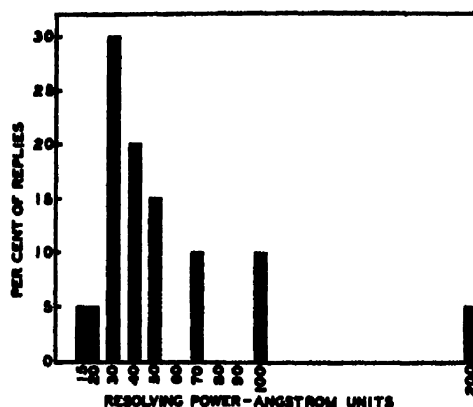


FIG. 1. A frequency diagram showing the resolving power which a number of individual electron microscopists considered to be the best they had obtained with their own instruments. The data were obtained from replies to a questionnaire sent to all members of the EMSA by the Committee on Resolution.

with the type of specimen studied may be considerably poorer, say 100A. In such cases, where-ever possible, the resolution present in the original prints should be stated.

The Committee also suggests that the following distinctions be made in discussions on this subject in order to denote more definitely the type of resolving power or resolution which one is describing.

1. The *QUALITY* of a micrograph.
2. The *RESOLUTION* of a micrograph.
3. The *INSTRUMENTAL RESOLVING POWER* of a specific electron microscope.
4. The *CALCULATED LIMIT OF RESOLVING POWER* of an actual lens system.
5. The *THEORETICAL LIMIT OF RESOLVING POWER* of an electron beam.

These five divisions are described further as follows:

1. **Quality.** Basically the quality of a micrograph is determined by whether or not the picture gives the complete morphological information about the specimen. In a more practical sense, quality can be considered as a measure of the extent to which the micrograph provides the information which was sought in taking it. The Committee wishes only to point out here that quality and resolution are not synonymous but rather that resolution is but one factor contributing to quality. Many other factors, e.g., the population density and dispersion of the particles of the specimen, the absence of dirt,

etc., from the photographic plate, the lack of all artifacts, will likewise contribute to quality. Many of these will be quite independent of and have no effect on the resolution in the picture.

2. Resolution of a Micrograph. As already stated, the resolution in a micrograph is related to quality in that it sets a definite lower limit to the detail of structure which can be seen. However, there are other reasons for which one may be interested in resolution in a micrograph, particularly in its quantitative evaluation. Examples of such are:

a. To determine the minimum error in measurement of a size made on the micrograph particularly in determinations of particle size distributions where the sizes are near the resolution present.

b. In evaluating the effects upon resolution of various methods of specimen preparation and instrument adjustment.

c. In determining the extent to which physical properties of the specimen, density, thickness, etc., limit resolution.

d. In determining the extent to which the capabilities of the instrument have been utilized in a particular micrograph and as a means of detecting departure from the ideal performance of the instrument.

In general, the resolution in a micrograph is of interest either as a measure of the limiting detail available or as an integrated quantitative evaluation of the quality of the operation of the instrument.

In this connection, it should be pointed out that the resolution present will vary from point to point in a single micrograph, a fact which has been verified by measurements made by some of the Committee members and which is undoubtedly familiar to many electron microscopists. The phenomenon is probably due to the variation in the physical properties of the particles at those points, to different levels in the specimen, or to aberrations of the lens. It is brought out here merely to indicate the necessity for caution in attempting to measure the resolution in a single micrograph.

3. Instrumental Resolving Power. With this division we arrive at the categories concerning resolving power. Here we are considering capabilities, theoretical or actual, rather than the extent to which such capabilities have been realized in a particular application.

A prime but not sole factor determining the resolution in a micrograph is, of course, the

resolving ability of the microscope used. This ability may be called the instrumental resolving power and it may be defined as the best resolution obtained with a given microscope when all conditions of specimen, photographic, and operational factors are ideal. It is, therefore, simply the best that instrument can do. It is significantly different from the previous category of resolution in that the choice of specimen is arbitrary in this determination.

A measurement of such instrumental resolving power would be of significance for several different purposes, for example:

a. To compare various designs, lens systems, etc.

b. To compare various microscopes of the same make or of different make.

c. To evaluate the effect of variations in different operating conditions such as apertures, angle of illumination, sources, etc.

d. To evaluate various photographic materials and processes.

From the results of answers to the questionnaire sent out to members of the Society a measurement of this type of resolving power is by far of greatest interest. Furthermore, it would probably be of advantage to have a standard method for measurement of this type of resolving power, particularly in comparisons between microscopes and microscopists. However, the Committee has not found a method which appears sufficiently accurate or sufficiently reproducible to recommend it for this purpose. It is possible, however, that experimental testing of some of the methods to be mentioned later will prove them suitable.

4 and 5. Calculated Limit of Resolving Power and Theoretical Limit of Resolving Power. The last two categories are described together because they have two points in common. First, both deal with limiting values obtained only mathematically and not in practice. Second, both are of considerably less interest to the Society as a whole although they are important to builders and manufacturers of microscopes.

The calculated limit of resolving power is that which one obtains for a specific instrument by the mathematical calculation of the effects of known limiting factors such as lens aberrations, electron velocity, etc. Whereas the instrumental resolving power is a measure of what an instru-

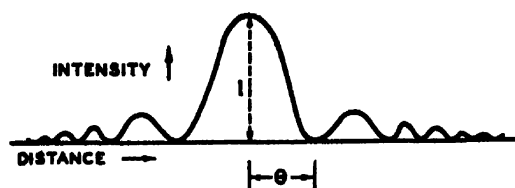


FIG. 2. The radial distribution of intensity in the image of a point source.

ment can do, the calculated limit of resolving power is a measure of what it should do according to theory. The ratio of these two is, therefore, a measurement of the efficiency of design of the instrument.

The theoretical limit of resolving power is merely the resolving ability of electron waves independent of any real instrument. It is commonly approximated by means of the Abbe formula, $d = \lambda/2N \cdot A$, or the Helmholtz or Rayleigh modifications of this. The validity of such approximations has been frequently questioned particularly in applying them to electron rather than light optics. However, a discussion of this point is believed beyond the scope of this report.

III. METHODS OF MEASUREMENT

From the experience of the Committee members, the published literature, and the answers to the questionnaire, it appears that there are seven possible types of measurements which may be made in connection with resolution or resolving power. These methods are briefly described at this time together with the Committee's evaluation of each.

1. Minimum Distance Between Particles. This method was indicated by nearly 50 percent of those giving positive replies to the question. "How do you measure resolution?" It is undoubtedly based upon the definition of resolution as a minimum distance between two points in the object whose images can be seen as separate. With this definition or with this method of measurement, the Committee has no disagreement. There are, however, several precautions to be exercised and several limitations of the method.

The first of these involves the translation in practice of the word *points* in the definition to the word *particles* in the method. It has been

shown by physical optics¹ that the distribution of energy about the image of a *point* source produced by a lens of circular aperture follows a curve similar to that shown in Fig. 2. In this curve the distance, θ , is a function of the lens system, its aperture and aberrations, while the height I is a function of the intensity of the light source. Now, if we consider two points gradually approaching each other as in Fig. 3, it will be seen that their images begin to overlap significantly before the points themselves coincide and at some stage the two images will not be visually separable. At just what separation this occurs is determined by individual physiological factors, but for our purposes here we may adopt the frequently used conclusion¹ that, for the normal eye, the images will be undistinguishable when the first minimum of one image reaches the central maximum of the other, i.e., when $d = \theta$. At this stage the actual separation of the points in the object (d /magnification) is the resolving power of the system.

However, let us now consider the relations which exist in the images of actual particles. If, with the same optical system as before, we should determine the intensity distribution about an image of a self-luminous particle of diameter D and in a plane section along a diameter, we would find that the curve for such an image would be similar to that of the image of a point with one important exception. In the case of particles whose diameter, D , is of the same order of magnitude as the resolving power or longer, the θ_1 for the particle image will be determined by that particle size as well as by the resolving power for a point. When, therefore, one measures the minimum distance of approach of two such particles, the resolution obtained will also be determined by the size of such particles if the measurements are made from center to center in the images.

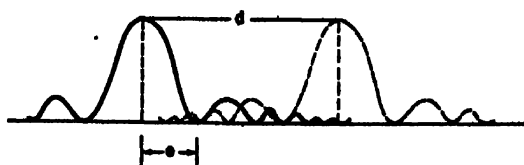


FIG. 3.

¹ John K. Robertson, *Introduction to Physical Optics* (D. Van Nostrand Company, New York, 1929), p. 233.

It, therefore, follows that the true resolving power of the system will be obtained only when the particles which are used as points are considerably smaller than the resolving power of the system. For electron microscopes this means that particles sufficiently small to behave as points must be only a few angstrom units in diameter or nearly of atomic dimensions. Some considerations regarding contrast of such particles toward electrons indicate that it may not be possible to see such particles, particularly those composed of low atomic number elements, even though the ability for linear resolution of the distance between them is available.

The significance of these considerations for this method of measuring resolving power lies in the fact that it shows that the measurement is always conservative for real particles of appreciable size, and, furthermore, that it requires particles of a size less than the resolving power being measured and of high contrast, *vis.*, of high atomic weight elements. For this reason it is suggested that the method be used to determine instrumental resolving power only when a suitable specimen is imaged, and that the specimen for that purpose be heavy metal particles of sufficiently fine size. The following specimens appear possible of application for such a purpose:

- a. Evaporated metal films where the particles do not tend to coalesce or agglomerate. Evaporated films of 5-10A thick chromium appear excellent for this purpose.
- b. Colloidal gold, silver, or platinum. Argyrol in 1:1000 dilution with water is fair although other materials are frequently present which obscure the picture.

As a test of the effect of different variables on the instrumental resolving power, the method suffers from the fact that one can never be certain that there is present somewhere in the field a pair of particles which are at the requisite minimum distance apart. For this reason, too, the method would appear to give conservative values of resolving power. Certainly micrographs of two different fields taken with two different microscopes cannot be quantitatively compared by such measurement except in a most approximate fashion. However, if it is possible to obtain a field in which the separation of uniformly small particles varies over wide values about the separation which can just be resolved, such a field

could be used to ascertain the effects of variations in operating conditions.

Obviously a distinct advantage of this method is its convenience since it requires no auxiliary measuring device other than a ruled scale.

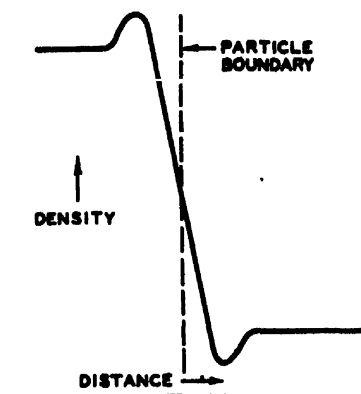


FIG. 4. A qualitative representation of the densitometric trace which is obtained across the photographic record of an electron microscope image.

2. Minimum Particle Diameter. It has been suggested that the diameter of the minimum sized particle present in the field be used as a measure of resolving power. As previously indicated it should be possible in theory to obtain images of particles of increasingly small diameter until the size of the image is no longer affected by the size of the particle. Such an image size should bear a distinct relation to the resolving power of the optical system. However, here again the important practical consideration of contrast must be taken into account. The Committee cannot state at the present time whether such a method would be of practical attainment—experimental tests should be carried out. It can be said, however, that this method suffers from the same disadvantages as that of minimum particle distance.

3. Contour Sharpness. If an in-focus image of a sharp and reasonably straight edge of a particle be recorded photographically in the electron microscope, a densitometer trace perpendicular to this edge appears as shown in Fig. 4. On such a curve one may measure a variety of constants which will bear some relation to the resolution present. Such measurements are as follows:

- a. The lateral distance from maximum to minimum translated by magnification factors to a size in the object.

- b. The slope of the straight line portion of the curve.
 c. Following the suggestion of von-Ardenne,³ the lateral distance, translated to the object, between points on the curve at 15 percent and 85 percent of the difference in intensity outside and inside the particle.

Such values are a direct measure of what one might call the sharpness of the image. This in turn is undoubtedly a direct function of the resolution present though just what function the Committee cannot say.

A significant advantage of this method lies in its reproducibility, particularly for different fields of the same specimen. It appears to be somewhat sensitive to changes in operating conditions also. It could, therefore, be used as a measure, though not a determination of the resolution in a micrograph or as a test procedure to study the effect of variations in operating conditions or of specimen preparation. Should it be possible to relate mathematically one of the values mentioned or a similar constant of the densitometric curve to the resolution present, it would be a convenient method of determining instrumental resolving power.

An interesting variation of this method is as follows: It is possible to predict from known values of lens aberrations and electron accelerating voltages the contour trace one should obtain in theory. Comparisons of this theoretical curve with that actually obtained gives a measure of the efficiency of the system. Such a test could be of interest in lens and instrument design.

4. Rounding of Particle Corners. A variation of the contour width measurement is found in the optical rounding of images of sharp corners

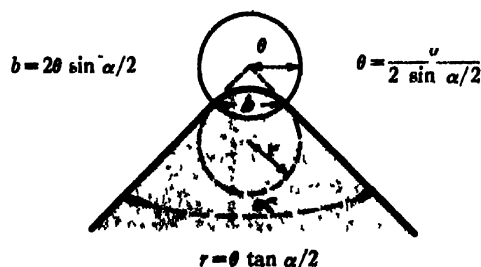


FIG. 5. A diagram illustrating the measurements used to determine the resolution in an electron micrograph from the apparent rounding of sharp corners.

³ Manfred von Ardenne, *Elektronen-Übermikroskopie*. Reproduced under Authority of Alien Property Custodian (Edwards Brothers, Inc., Ann Arbor, Michigan, 1943), p. 279.

of particles. This method is a development from the suggestion of Kausche and von Borries⁴ as to the cause of such rounding. Essentially these authors assume that such a corner behaves as the ideal point and that, at the place where the circle of the first diffraction disk of that point intersects the straight edges of the particle, the deviation of the image from the true edge begins. This is shown in Fig. 5. The measurements which may be made thereon are indicated. Notice that these measurements are a function of the angle between the intersecting edges and such should be taken into account in any determination.

In actual practice the extent of rounding appears to be a more complex phenomena than the simple assumption of Kausche and von Borries. Determination of the distance, θ , are considerably larger than minimum particle distances. It seems quite likely that such rounding is more closely associated with the contour width and perhaps bears the same relation to resolution. More experimental work on this subject appears necessary. It is also known that contamination in the instrument due to the electron bombardment rounds the corners.

Such a method would prove of the same use as the contour width measurement. It has the advantage of not requiring a densitometer but the disadvantage of being less precise.

5. Rounding of Particle Intersections. A serious disadvantage of the method of measurements of rounded corners lies in the fact that one cannot be certain that the corners are not in physical reality rounded themselves, perhaps even by the bombardment of the electron beam. To avoid this a method has been suggested by Dr. Sergei Feitelberg⁴ of the Mt. Sinai Hospital in which the rounding is measured not at a particle corner but at the intersection of the straight edges of two particles. There should at least in theory be no reason for more matter to be knocked off at such an intersection than along the remainder of the edge.

It would appear that otherwise the same statements apply to this method as given for the rounding of particle corners. The Committee

⁴ B. von Borries and G. A. Kausche, *Kolloid Zeits.* 90, 132-141 (1940).

⁴ Sergei Feitelberg, private communication.

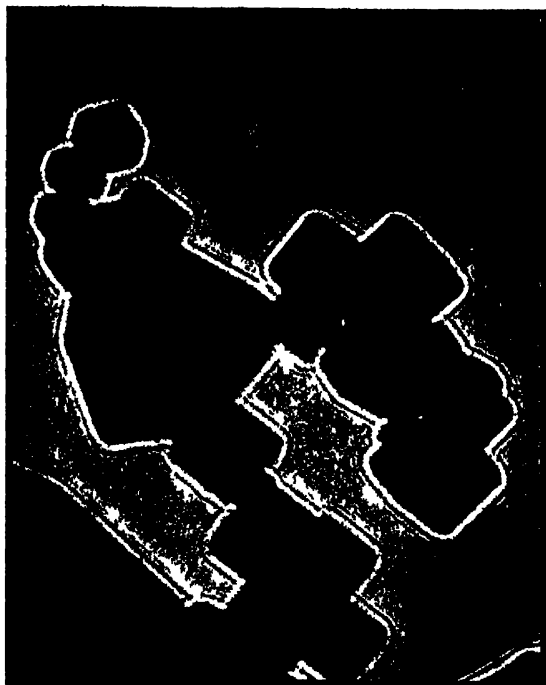


FIG. 6a. An out-of-focus electron micrograph of magnesium oxide smoke showing Fresnel fringes.

would recommend the additional precautions that the two edges be as near as possible in the same focal plane and that there be no deposition of other materials at the intersection.

6. Quality Estimation. A method was recently described by Hillier⁵ in which micrographs are visually graded for quality and this grade is expressed in terms of the maximum final magnification one estimates the picture quality will allow. Such a method has proven of value to him in evaluating over-all level of performance and progress in that direction. In such a method one examines the negative at about 10–20 times magnification. Quality may be judged by whatever factors are of particular interest—amount of contrast, accuracy of focus, symmetry of optical fringes or borders, exposure, distortion, and many others. Such a method could be useful for approximating “batting averages.”

7. Measurements from Fresnel Fringes. A method of considerable promise for measuring instrumental resolving power is the measurement of Fresnel diffraction fringes in an out-of-focus micrograph of a straight edge of a particle. This

⁵ J. Hillier, *J. App. Phys.* 16, 469 (1945).

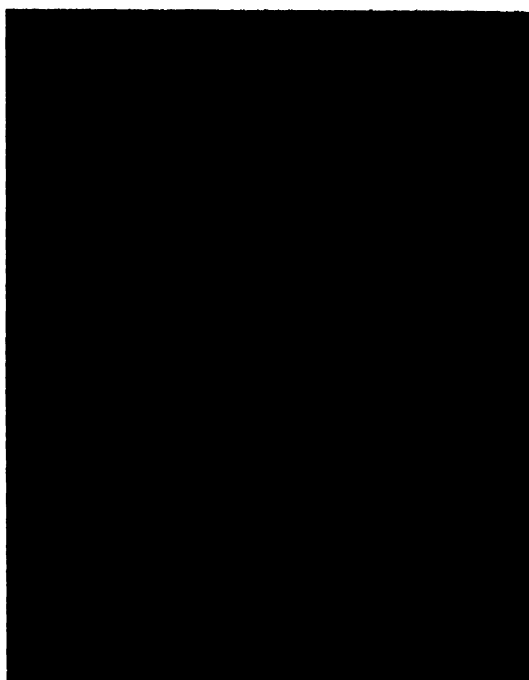


FIG. 6b. An electron micrograph of part of the same field of view as in (6a) which is only slightly out of focus and shows closely spaced Fresnel fringes.

method, suggested by Ruska,⁶ and Hillier,⁷ is an indirect result of work done by Boersch⁸ though other authors have written on the subject. Images of such fringes are shown in Fig. 6. A densitometric trace perpendicular to such fringes is shown by Hillier.⁷ The method consists briefly of measuring on such a trace the lateral distance between the maximum and minimum which can just be resolved. This distance, translated to the object plane by magnification factors, gives a value for instrumental resolving power.

To obtain such fringes in an electron microscope image requires the following conditions: First, the electron source must be as small as possible. This may be achieved either by using an extremely small condenser aperture—Boersch used a 5-micron aperture—or by obtaining a very small source free of secondary scattering.

The condenser current should be near maximum and the objective current should be somewhat lower than the in-focus value, i.e., the objective should be focused on a plane somewhat

⁶ E. Ruska, *Kolloid Zeits.* 105, 43–52 (1943).

⁷ J. Hillier, *J. App. Phys.* 17, 307 (1946).

⁸ H. Boersch, *Physik. Zeits.* 44, 202 (July 1943).

beyond the specimen in the direction of the condenser.

The phenomenon of Fresnel diffraction provides as the test for resolving power a series of secondary electron sources rather than physical particles. As one goes outward from the diffracting edge, the maxima and minima become increasingly closer both in angular direction and in intensity difference. It is not certain at the present time whether the limiting pair which are just visible are determined by their angular spacing and the resolving power of the microscope, or by the intensity difference and the contrast obtained therefrom on the photographic plate. Experimental tests might show this. Nevertheless, the method appears to be of considerable promise for testing the instrumental resolving power. While the out-of-focus condition of the objective changes the magnification somewhat from the in-focus value, an exact evaluation of the change can be made by a direct comparison of the in-focus and out-of-focus images.

IV. SUMMARY

In summary the Committee presents five quantities which are associated with resolving

power and seven methods of measurement for such quantities. These are as follows:

1. The quality of a micrograph—probably best evaluated by visual inspection for the factor or factors of interest.
2. The resolution of a micrograph—there is no known method of obtaining this value for all micrographs. Probably the indirect measures of contour sharpness, or of corner or intersection rounding would serve this purpose best.
3. Instrumental resolving power—this value is probably best determined by measurements of Fresnel fringes. The minimum particle distance or the contour width may also be used.
- 4 and 5. Calculated and theoretical limits of resolving power—these values are mathematically evaluated and the computations involved are beyond the scope of this report.

As indicated several times, much experimental work seems desirable in this field. Until such methods have been thoroughly tested for preciseness and accuracy, the Committee cannot recommend that the Society standardize upon any one.

W. G. KINSINGER (*Chairman*)

J. HILLIER

R. G. PICARD

H. W. ZIELER

Electron Microscope Examination of the Microphysical Properties of the Polymer Cuprene

JOHN H. L. WATSON AND K. KAUFMANN

Plant Research Department, Shawinigan Chemicals Limited, Shawinigan Falls, Quebec

(Received July 10, 1946)

The microphysical structure of cuprene is described from electron micrographs and some suggestions are made concerning the mechanism of the formation of this substance. Cuprene samples prepared by polymerization of acetylene in the presence of finely divided cuprous oxide are shown to be composed of hollow fibers characterized by both longitudinal and transverse structure. The striking appearance in electron micrographs of this type of cuprene serves to identify it in clogging deposits taken from acetylene lines. This is illustrated by a deposit taken from a reactor used in the process of hydrogenation of acetylene to ethylene. Cuprene specimens formed by alpha-ray bombardment of acetylene gas are shown to be composed of round particles joined by short, straight, thick necks which are about 500 to 1000A in length. These are samples of cuprene which have been bombarded further by alpha-particles after their initial formation. An oxidized sample has round particles of mean diameter 4900A; an unoxidized sample has a mean diameter of 3800A for the round particles. There is evidence to suggest that alpha-ray cuprene is liquid in nature. The two types of cuprene are so different physically that a change in nomenclature is in order.

CUPRENE, PRODUCED CHEMICALLY IN A CATALYZED PROCESS

THE first references to cuprene are by Erdmann and Koethner.¹ It has been

studied since by numerous investigators and comprehensive summaries are given by Calhoun²

¹ Erdmann and Koethner, *Zeits. f. anorg. Chemie* 18, 48 (1898).

² Calhoun, *Can. J. Research* 15, 208 (1937).



FIG. 1. Brown cuprene made chemically between 200 and 240°C with a cuprous oxide catalyst
Note longitudinal and transverse structure in the fibers $\times 20,000$

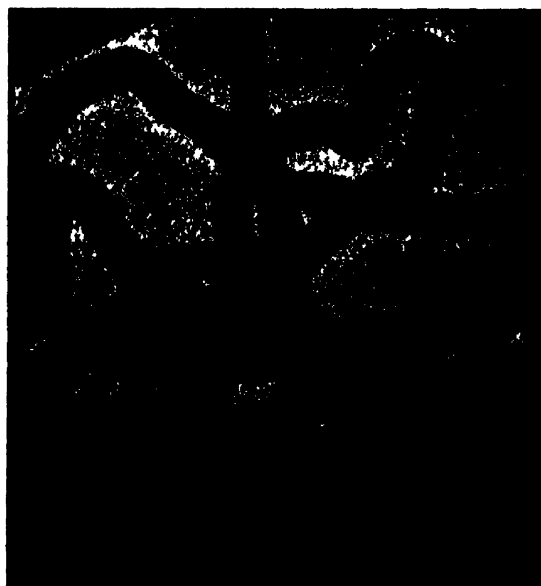


FIG. 2. An enlarged section of Fig. 1 to show longitudinal structure more clearly. $\times 36,000$.

and others. It is a solid, amorphous material varying in color from a dark brown to a light yellow, and in physical appearance to the eye from a light, flocculent powder to a relatively dense, cork-like mass according to the conditions of preparation. Chemically, its nature is uncertain but it appears to be a polymer of acetylene with formula $(C_nH_m)_x$.

■ If it is prepared in the presence of copper or finely divided cuprous oxide as a catalyst, the polymerization starts at about 200°C, proceeds rapidly, and slows down gradually. Increase of temperature hastens the reaction but above 300°C carbon forms instead of cuprene. Solids resembling cuprene have been obtained also using tin, nickel, and cobalt as catalysts, and by the action of silent electric discharges, ultra-violet radiation, cathode rays, and alpha-particles upon acetylene gas.

The samples studied here were prepared by chemical means using finely divided cuprous

oxide as a catalyst between temperatures of 200°C and 240°C.

The cuprene presents a very striking appearance in electron micrographs and was photographed first by J. H. L. Watson and L. T. Newman with the electron microscope at the University of Toronto.³ These pictures showed it to be composed of fibers which, from intensity characteristics in the micrographs, appeared to be tubular.

This cuprene is mounted for the electron microscope by rubbing the corky substance across a specimen holder with a Formvar film attached or across a filmless holder. Original magnifications of $\times 5000$ are sufficient to show up detail existing within the fibers.

The fibrous, tubular nature of this cuprene formed chemically with a cuprous oxide catalyst is observed in both flat and stereoscopic electron micrographs. In addition, both a *transverse* (across the fiber) and a *longitudinal* (along the



FIG. 3. Brown cuprene made chemically with cuprous oxide catalyst. Note structure and loosened skin on fiber. $\times 20,000$.

³ Burton and Kohl, *The Electron Microscope* (Reinhold Publishing Corporation, New York, 1946), second edition, p. 267.

fiber) structure are observed (Figs. 1–3). Figure 2 is an enlarged section of Fig. 1.

Considering first the *longitudinal* structure, two types are observed. From Fig. 2 and from certain stereoscopic views it is concluded that most of the larger fibers are composed of several finer ones lying parallel to each other and touching in a bundle. This is evident especially in those where the thickness is such that the electrons are able to penetrate with relatively small scattering. The finest longitudinal structures have diameters of the order of 80A. This is presumably the order of the diameter of the finest fibers and is equivalent to 50 to 60 carbon bond distances.⁴

There is great variety in both fiber shape and fiber diameter, but the tubular nature seems to be universal. For this reason a calculation of the approximate number of carbon atoms in the fiber circumference is in order. Again, if a fiber of circular cross section is considered with the minimum diameter 80A, it possesses 164 or 180 carbon bond distances in the length of its circumference, according as single or double bonding is assumed between the atoms.

Intensity variations over the micrographs signify either a true tube as the nature of the fibers, with carbon atoms clustered only at the skin of the tube; or merely that there is a *greater* concentration of carbon atoms at the skin than in the central regions. From stereoscopic views in general, as represented by Fig. 4 (location marked by the arrow), the tubes are seen to be hollow and are consequently true tubes with the atoms of the substance clustered at the edges. This stereoscopic pair is chosen as Fig. 4 because the tube marked is particularly interesting. It is not only hollow, and thin-skinned, but *hexagonal* in cross section. In addition it is noted that from the corners of the hexagon dark linear areas proceed down the fiber. These linear areas of apparently denser or thicker nature have been observed in other fibers of various shapes and contribute to a second type of longitudinal structure. They may account also for some of the longitudinal structure observed in Figs. 1 and 2.

⁴ Values for single bond and double bond distances are 1.54A and 1.38A, respectively. Daniels: *Chemical Kinetics* (Cornell University Press, Ithaca, New York, 1938), p. 226.



FIG. 4. A stereoscopic view of brown cuprene made with cuprous oxide catalyst. Note the hollow fiber of hexagonal cross section at $\times 8000$.



FIG. 5. A stereoscopic view of brown cuprene made with cuprous oxide catalyst. Note the ribbon-like, spiraled fiber. $\times 3800$.

Cross-sectional areas of circular, oval, square, rectangular, ribbon-like, and hexagonal ap-

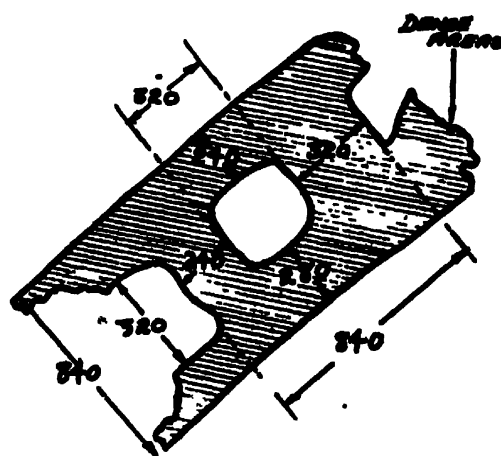


FIG. 6. Drawing to give the dimensions of the dense and light areas in a typical piece of cuprene possessing transverse structure. All dimensions are in angstrom units.

pearance have been recognized. Figure 5 is a stereoscopic pair of a ribbon-like fiber. It is suggested that the polymerizing fiber takes the cross-sectional shape of the particular small area or particle of catalyst upon which it begins to grow. There may be some information to be gained here concerning active centers on catalytic surfaces; a combination of active center properties and molecule sizes possibly may determine both size and shape of the fiber cross section.

Perhaps more interesting than the longitudinal structure is the *transverse* structure evident throughout the micrographs. Typical occurrences are seen in Figs. 1 and 3 in fairly large fibers. These fibers appear to be divided by dense areas (recognized by dark regions in the micrographs) into cells, where presumably, again, the atoms of the substance are tending to cluster. Usually the length of one of these cells along a



FIG. 7. An enlarged section of Fig. 1 showing a fine fiber of diameter 170A which possesses transverse structure. $\times 65,000$.

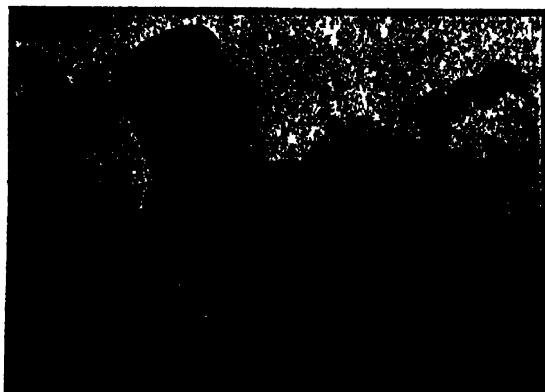
fiber is of the same order numerically as the value of the diameter of the particular fiber. At the mark of the arrow in Fig. 1 the transverse structure has dimensions approximating the values marked in the drawing in Fig. 6.



FIG. 9. A sample of cuprene taken from a German reactor used in the hydrogenation of acetylene to ethylene with a palladium catalyst on silica gel. $\times 20,000$



(A)



(B)

FIGS. 8A and B. Small filaments of newly formed cuprene grown upon an electron microscope specimen screen, showing small pieces of catalyst being carried along with the growing fiber of polymerizing cuprene. $\times 20,000$.

From a few shadow cast specimens studied to date the observations of structure noted here have been substantiated but there is evidence to suggest that the structure is even more complicated in nature.

A skin is observed on the fibers in many micrographs as in Fig. 3 where it is apparently becoming loosened and is peeling off. This skin is observed chiefly with larger fibers. In the very tiny individual ones as reproduced in Fig. 7 there is no visible skin, nor longitudinal structure, but a definite suggestion of transverse structure still persists. Figure 7 is a much enlarged section of Fig. 1.

Intensity variations over electron micrographs are not easily interpreted but in general relatively darker areas in an image indicate either thicker, or else physically or atomically denser, parts of the specimen. The dark areas here indicate parts of the cuprene fiber where clustering of carbon atoms occurs. There is another possibility that copper or cuprous oxide is able to migrate along the fiber and does so along the dark image areas described. The percentage of



FIG. 10. Alpha-particle cuprene oxidized and bombarded with further alpha-particles after the original formation. $\times 20,000$.

copper in cuprene is so small (3 percent) that it is most probable that clustering of carbon atoms in the preferred positions accounts for the structural pattern observed.

That metallic copper is actually associated with the fibers and is migrating along them, or is carried along with them, is supported by not very satisfactory transmission electron diffraction patterns of this cuprene where the only visible line was found corresponding to a grating space of $2.08\text{\AA}$.⁵ for the strongest line of metallic copper. Chemical tests also detect copper even in the very outermost parts of a cuprene growth far from the original catalyst. In the reducing atmosphere associated with cuprene formation it is likely that the cuprous oxide is reduced to metallic copper.

In order to study cuprene as it *first* begins to form it was grown upon copper specimen screens impregnated with cuprous oxide. It is found that

⁵ Hanawalt, Rinn, and Frevel, *Ind. Eng. Chem. Anal. Ed.* 10, 457 (1938).

the first forming cuprene begins its growth as a small spiraling fiber, as in Figs. 8A, and B. The spiral nature of the fibers persists in older samples. From observations of the newly formed, short fibers it is seen that they are carrying small pieces of some heavy material with them. We suggest that this is either cuprous oxide or metallic copper reduced from the oxide by hydrogen, and that it is carried along as the fiber grows, continuing in this way to catalyze the reaction. This mechanism would explain the presence of copper in the outer parts of the growth.

Some photographs were taken of cuprene, which had been raised in temperature slightly beyond 300°C . In addition to the usual fibers, small round particles of mean diameter $440\text{\AA}$ were detected. They were discrete and did not occur in chains although they are probably a type of acetylene black.

Cuprene is often a nuisance as it grows in acetylene lines. If cuprene is suspected to be the

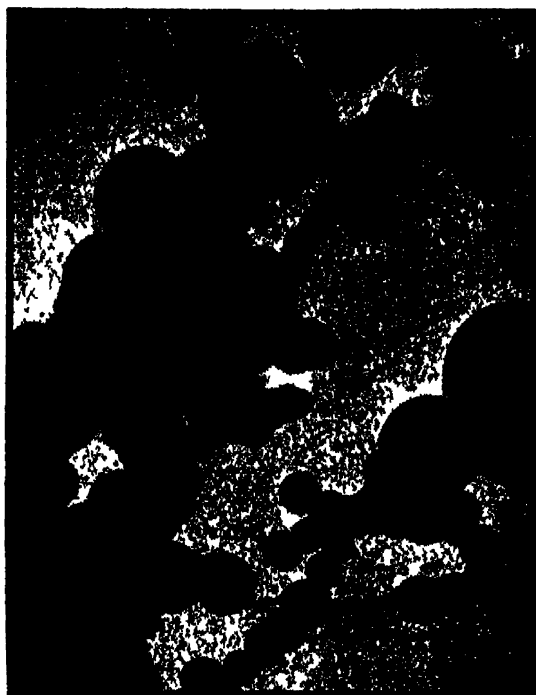


FIG 11 Alpha-particle cuprene unoxidized, and bombarded with further alpha-particles after the original formation. $\times 20,000$.

blocking agent in a line, it may be identified readily as such with the electron microscope because of its characteristic appearance when formed by a catalyzed chemical process. Figure 9 shows a deposit of a substance taken from a reactor which had been used in the process of hydrogenation of acetylene to ethylene at Anorgana, G m b H, Gendorf, Bavaria, with a palladium catalyst on silica gel. This was a hard, rocky deposit, quite unlike Shawinigan cuprene to the eye, and had to be ground and mounted from a water suspension for the electron microscope. From electron micrographs the deposit was found to contain chiefly cuprene mixed with some other unidentified materials. No transverse structure was observed in the fibers but they do possess some longitudinal structure. There are many thin, film-like fragments which are probably wall structure from the ground fibers.

ALPHA-PARTICLE CUPRENE

Cuprene may be formed also when acetylene gas is bombarded by certain atomic particles,



FIG 12 Chains of alpha-particle cuprene stretched to show the existence of joining necks $\times 20,000$.



FIG. 13. Part of the same field as in Fig. 12 after the chain had broken, showing permanent deformation of the stretched particles and necks of cuprene. $\times 20,000$.

such as alpha-particles from radon.⁶ At ordinary temperatures and pressures a light yellow powder is deposited from the process.

Two samples of alpha-particle cuprene which were made some twenty years ago were secured from Dr. S. C. Lind.⁷ That described as being "oxidized" was left open to the air for this time through a long, fine capillary and the probability is that it has oxidized slowly; alpha-ray cuprene adds about a quarter of its weight of oxygen in about three months time if it has free access to the air. The sample described as "unoxidized" was not open to the air. After their original formation both samples were subjected to further alpha-ray bombardment and probably lost some hydrogen in the process.

The physical appearance of this type of cuprene is distinctly different from that of cuprene formed by chemical means in the presence of a cuprous oxide catalyst. The alpha-ray cuprene is composed of comparatively large, round particles joined to each other by short, thick necks or rods of length approximating 500 to 1000Å. These necks appear to join the particles at right angles. Figure 10 shows the oxidized sample wherein the particles and joining necks are seen. The unoxidized specimen has the same general characteristics and is shown in Fig. 11.

The existence of these necks is illustrated in Fig. 12 where a broken Formvar film has stretched a chain apart and lengthened the necks without breaking them. Figure 13 shows part of the same field after one link in the chain has broken. The distorted particle shape remains. In Fig. 14 several stretched cuprene necks can be seen over a breaking Formvar film, and in addition it is seen that the "droplets" of cuprene have been distorted about the holes in the Formvar. The observations mentioned in this paragraph indicate that this cuprene has the properties of a very viscous *liquid*, and can be compared to rubber latex in some of its particle characteristics. This would explain why the cuprene gives no x-ray diffraction pattern.

⁶Lind, *The Chemical Effect of Alpha Particles and Electrons* (The Chemical Catalog Company, New York, 1928).

⁷Lind, Bardnell, and Perry, *J. Am. Chem. Soc.* **48**, 1556 (1926).



FIG. 14. Alpha-particle cuprene chains, showing stretching of the necks and deformation of the "droplets" about holes in the formvar film. $\times 20,000$.

In a recent article by Stanton and Henson⁸ a latex is described which possesses an appearance in electron micrographs very similar to that described here for the polymer cuprene. This property of large round particles joined by necks of the material would therefore appear to be common to some types of polymerized substances. On the other hand the fibrous structure of catalyzed cuprene is probably common to others, especially those grown upon a catalytic surface.

Little difference is immediately observable between the oxidized and unoxidized samples, but from particle size distribution studies of the "droplets," the mean diameter of the oxidized sample is found to be 4900Å and of the unoxidized sample to be 3800Å. This is a difference of 30 percent and is a significant difference for this type of counting procedure. The standard deviations for oxidized and unoxidized samples are 1300 and 2200Å, respectively, indicating a greater spread of particle sizes in the unoxidized

⁸Stanton and Henson, *Ind. Eng. Chem.* **18**, 468 (1946).

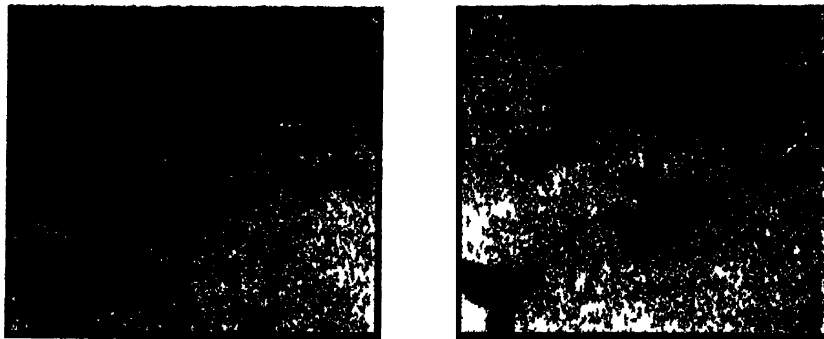


FIG. 15 A stereoscopic view of alpha-particle cuprene, oxidized, showing the "molecular model" appearance of the chains $\times 8000$

sample. The particle diameters range from about 3000 to 10,000A, in the oxidized sample, and from about 1000 to 13,000A, in the unoxidized material. In addition to these differences in particle size measurements it seems from stereoscopic views and from flat prints that the necks are shorter and consequently less frequently observable in the unoxidized cuprene. Figure 15 is a stereoscopic view of the oxidized alpha-ray cuprene, showing the "molecular model" appearance of the chains.

A few of the smooth, round particles have slight protuberances on the side (see arrows in Fig. 10). This is probably caused by the method of mounting the specimens which involves rubbing the cuprene out over a Formvar film. This procedure tears some particles apart and leaves a roughened section behind.

According to the "cluster-ion" hypothesis of Lind and Glocker^{9,10} some of the acetylene molecules are ionized by the bombardment process and these may either recombine to form smaller molecules or attract neutral molecules to form a cluster-ion. When the cluster-ion itself becomes finally neutralized a large molecule is formed. Lind⁷ assumes that at least 19 neutral molecules of acetylene are formed around each $C_2H_3^+$. It is very questionable that the round particles observed here are identical with Lind's clusters; at any rate many more than 19 neutral molecules of acetylene would be involved in clusters of the size of these particles.

⁹Lind and Glocker, J. Am. Chem. Soc. 53, 3355 (1931).

¹⁰Glocker and Lind, *The Electro-Chemistry of Gases and*

Other Dielectrics (John Wiley & Sons, New York, 1939).

If Lind's theory is correct, then one of the clusters might be the initial point of condensation about which other clusters could gather to form the large particles shown here. Since the condensation will occur in an unrestricted gas space, the product might be expected to occur in spherical form. The limiting size of these particles would be of the order of 13,000A, i.e., the largest

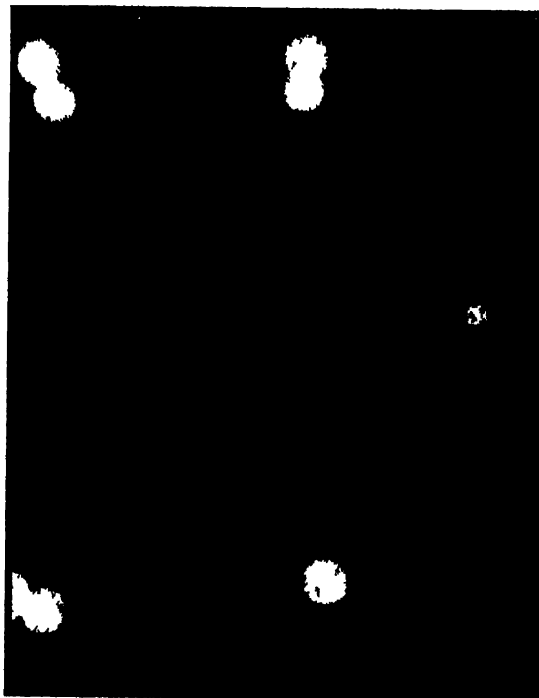


FIG. 16. A shadow cast specimen preparation of alpha-particle cuprene made with a filament height to specimen distance ratio of 1:4 with approximately 100A coating of chromium metal. $\times 10,000$.

particle diameter measured in these investigations.

The appearance of alpha-ray cuprene at the moment of its first formation could be determined by placing prepared electron microscope specimen grids in the polymerization chamber so as to catch the original droplets for examination. Direct evidence might then be made available concerning the mechanism of this polymerization phenomenon.

Cuprene samples were shadow cast with a ratio of filament height to specimen distance of 1:4. Measurements made on the plates gave an average ratio of 1:3 for particle diameter to shadow length, signifying that, although the cuprene "droplets" appear to be spheres in the usual micrographs, in reality they are flattened somewhat as they lie on the Formvar film. This fact was not observable in stereoscopic pairs, and is further evidence that this material is a viscous liquid. Figure 16 is a shadow cast preparation of alpha-ray cuprene coated with approximately 100A of chromium metal at the angle described above.

No structure is visible within the alpha-ray cuprene and its appearance and other properties are so different from that formed chemically that it would probably be worth while to change the

nomenclature and call it by a different name. Previously it had been supposed that all "cuprene" had the same physical structure. In this paper are three samples wherein differences are found. The Shawinigan cuprene is similar to the German sample in that each is fibrous, but the structure is different. Alpha-ray cuprene again is different to both.

From these results we may conclude that any "cuprene" made by a bombardment process (electrons, alpha-rays, photons, etc.) will possess a round-particle structure of some type; the polymerization process occurring in a discrete volume about the point of impact of the bombarding particle and a molecule. Whether the connecting necks will always occur is not known. On the other hand any catalyzed formation produced chemically on a surface will grow into a *fiber* of cuprene.

ACKNOWLEDGMENT

The authors wish to thank Dr. S. C. Lind, Dean of the Institute of Technology, University of Minnesota for the samples of alpha-ray cuprene, and Dr. Robley C. Williams of the Physics Staff of the University of Michigan for shadow casting several specimens.

Calendar of Meetings

December, 1946

- 17 Institute of Aeronautical Sciences, New York, New York
- 26-28 Geological Society of America, Chicago, Illinois
- 26-31 American Association for the Advancement of Science, Boston
Massachusetts

January, 1947

- 3-4 American Physical Society, Los Angeles, California
- 15-17 American Society of Civil Engineers, New York, New York
- 23-25 American Physical Society, New York, New York

February, 1947

- 20-22 Optical Society of America, New York, New York
- 24-28 American Society for Testing Materials, Philadelphia, Pennsylvania

May, 1947

- 1-3 American Physical Society, Washington, D. C.

June, 1947

- 16-20 American Society for Testing Materials, Atlantic City, New Jersey

A Machine for the Application of Sand in Making Fourier Projections of Crystal Structures*

DAN McLACHLAN, JR. AND E. F. CHAMPAYGNE

Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut

(Received June 1, 1946)

This paper presents a method of using sand in building models of two-dimensional Fourier projections in crystal structure analysis. The machine for producing these models is described and the results are illustrated by projections of known structures. The mathematical principle of this method of summing Fourier series is identical to that which Bragg developed in England. The chief difference is in the use of sand instead of photographic film as the medium for producing contrasts in electronic density throughout the projection. The method herein described has a number of features which contribute to its usefulness: A two-dimensional Fourier projection comprising the addition of one hundred and fifty terms may be made in about four hours by a single worker. The time required is not determined by the size of the unit cell. The subdivisions of the projection face are sufficiently fine to assure an accuracy as great as the x-ray data from copper radiation will warrant when a scale of one angstrom to two centimeters is used. Upon embarking on a new structure problem, no changes in the equipment are necessary; as the initial set of templates suffices for all future work. The minimum interplanar spacing that the machine will accommodate is determined by the characteristics of copper radiation rather than the order (h , k , or l) by which the spacings are characterized. The resulting sand model of the projection is the same shape as the unit cell projection with regard to both the angle between the edges and the ratios of the edges of the cell. When the x-ray data warrant, a distinction can be made between heavy atoms and lighter atoms by the differences in the height of the sand. Photographs of the projections (see Figs 7 and 9) portray atoms as distinct spots.

THE GENERAL CONCEPT OF WAVE ADDITION

WHEN a diffraction pattern of a single crystal is taken on any of the moving film types of camera, each diffraction spot presents three measurable parameters: two of position and one of intensity. When these measurements are used in making a projection of the crystal on one of its faces, the measured values are introduced into a Fourier series of the type

$$\rho_{xy}' = \sum_h \sum_k |F_{hko}| \cos [2\pi(hx + ky) - \alpha_{hko}], \quad (1)$$

in which there are as many terms as there are diffraction spots. Each term in the series

$$(S_{xy})_{hko} = |F_{hko}| \cos [2\pi(hx + ky) - \alpha_{hko}] \quad (2)$$

requires for its application the introduction of four measured parameters, h , k , F_{hko} , and α_{hko} . Of these four parameters, two, h and k , may be deduced from the two measurements of position of the diffraction spot on the film, and F_{hko} is directly related to the measured intensity, while

the fourth term, α_{hko} , must be obtained by other means.

If Eq. (2) is plotted out as a two-dimensional graph, a surface is thereby produced whose shape resembles a piece of corrugated sheet iron as shown in Fig. 1. If a sinusoidal sheet of material, $ABCD$, passes through the imaginary plane, $EFGH$, the distance S (of any point at xy on the

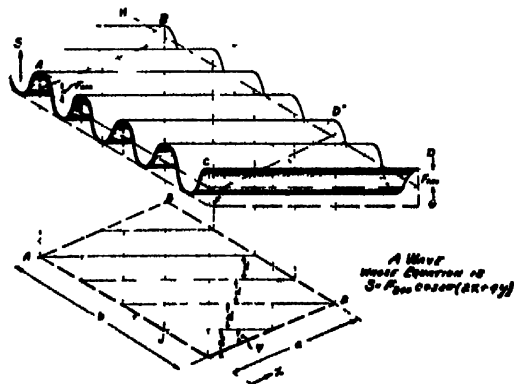


FIG. 1. The wave associated with the plane having Miller indices $hko=240$. The projection of the wave crests on the plane $A'B'D'$ shows the meaning of d and ψ . In this illustration $\alpha=0$.

* Presented in part before the Crystallographic Society in Northampton, Massachusetts, on March 23, 1946.

corrugated sheet) above or below the imaginary plane is expressed mathematically in terms of the coordinates x and y in the plane by Eq (2). More specifically, Fig. 1 shows the corrugated sheet oriented in such a position with respect to the reference plane that the wave crests cut the x axis twice in the unit distance (a) and cut the y axis four times in the distance (b). This means, by definition, that h is 2 and k is 4. In the figure the amplitude F_{hko} is arbitrary and the phase α is zero. The expression

$$(S_{xy})_{240} = F_{240} \cos 2\pi(2x+4y) \quad (3)$$

is the actual equation of the corrugated sheet shown in in Fig. 1 in terms of x and y and is applicable whether or not the x and y axes are perpendicular, as long as a and b are the unit distances. It is convenient that the perpendicular distance d_{hko} between the crests is equal to the interplanar distance found from the x-ray data

$$d_{hko} = n\lambda/2 \sin \theta, \quad (4)$$

and the angle ψ that the wave crests make with

the x axis is

$$\psi_{hko} = \arcsin (hd_{hko}/a). \quad (5)$$

This all means that for any crystal whose unit cell dimensions in the x and y directions are a

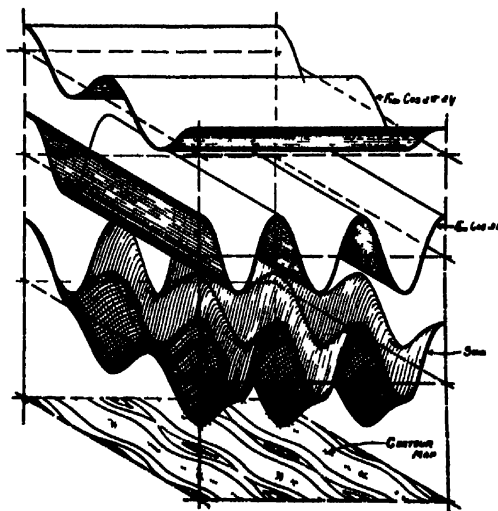


FIG. 3 The physical meaning of the addition of two waves and the relationship of their sum to a contour map.

and b , respectively, a wave of the type shown in Fig. 1 may be imagined as associated with every diffraction spot. The problem is to find a convenient method of adding together as many of these waves as there are diffraction spots in the x-ray pattern. Figure 2 shows a number of such waves in position for addition. The physical meaning of wave addition is illustrated in Fig. 3 where two simple waves corresponding to diffraction spots having Miller indices $hko=020$ and 300 , respectively, are added to form a warped sheet; and the contour map representing the warped sheet is projected on the base. The equation of the top wave is

$$S_{020} = 2 \cos 2\pi 2y \quad (6)$$

and of the second wave is

$$S_{300} = 4 \cos 2\pi 3x, \quad (7)$$

while the warped sheet obeys the equation

$$\rho_{xy}' = S_{020} + S_{300} = 2 \cos 2\pi 2y + 4 \cos 2\pi 3x. \quad (8)$$

In the contour map dotted lines represent "valleys" while solid lines represent "hills" on

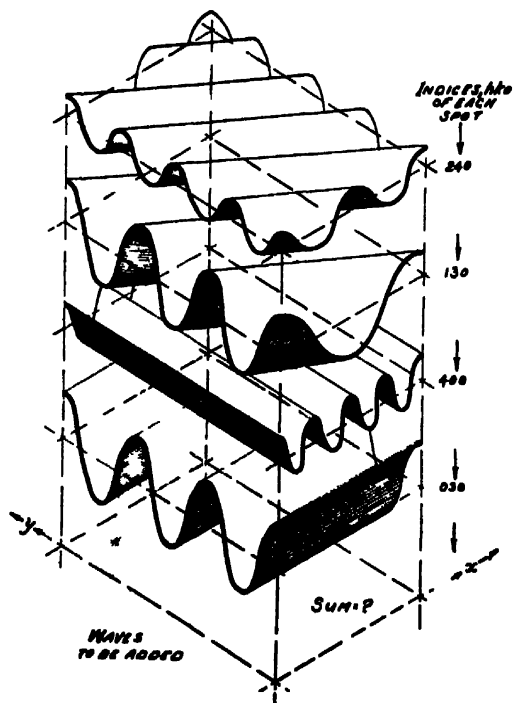


FIG. 2. A number of waves in position for adding on the basal xy plane.

the warped sheet. Such a contour map is the most common way of representing molecular structures projected on a plane, such maps being obtained from the analytical summation of the data from many dozens of x-ray diffraction spots.

THE BRAGG METHOD

W. L. Bragg¹ conceived a novel way of adding the waves associated with the terms in a Fourier series by using transparent photographic films. All of these films had one thing in common, that is, each film had on it a photographic impression consisting of parallel dark and light bands, so processed that the variation in the photographic density of the film in a direction perpendicular to the bands followed a modified sine law. Each film had, however, two characteristics which were unique: (a) the perpendicular distance between the parallel bands on any given film was proportional to d_{hko} , the interplanar spacing

characteristic of a diffraction spot; and (b) the direction of the parallel bands corresponded to the direction that an hko family of planes cuts the xy face of the unit cell. It may be seen that any one of the Bragg strips has all the characteristics of a wave such as is shown in Fig. 1. The summation as performed by Bragg was accomplished by preparing films which had the spacing d_{hko} and direction ψ_{hko} governed by the position of the x-ray diffraction spots; then parallel light was shone through each film in succession upon a photographic plate. The sum of all these registered photographic impressions upon a single plate constituted a projection whose density depicted the molecular structure of the compound from which the x-ray data was obtained.

The amplitude, F_{hko} , in each of Bragg's films was incorporated in the time of exposure of the plate through each film.

THE SAND MACHINE

The machine herein described is based on the thought that any one of a great variety of substances might be forced to play the role that the photographic emulsion plays in the Bragg method. The choice of an alternate method is determined by (a) the ease by which a machine may be developed, (b) the convenience and speed in operation while projecting a particular structure, (c) the absence of extraneous labor in changing from one problem to another, (d) the accuracy of a chosen method, and (e) the need for a method which produces a projection to scale and which does not require distortion of the projection before interpretation. One of the most important decisions had to do with the choice of the substance to be distributed over the projection face to simulate electronic density. A substance which was found to be satisfactory was 100-140 mesh Ottawa sand. This material has a low angle of repose, is not affected by moisture, does not take on a noticeable electrostatic charge, is resilient on the bounce, and has other properties which add up to assuring reasonable duplication in flow measurements from rollers and screens.

Figure 4 shows schematically the principle of the sand machine. Sand, S , is loaded in a hopper, H , where it rests on a roller, R , which forms the

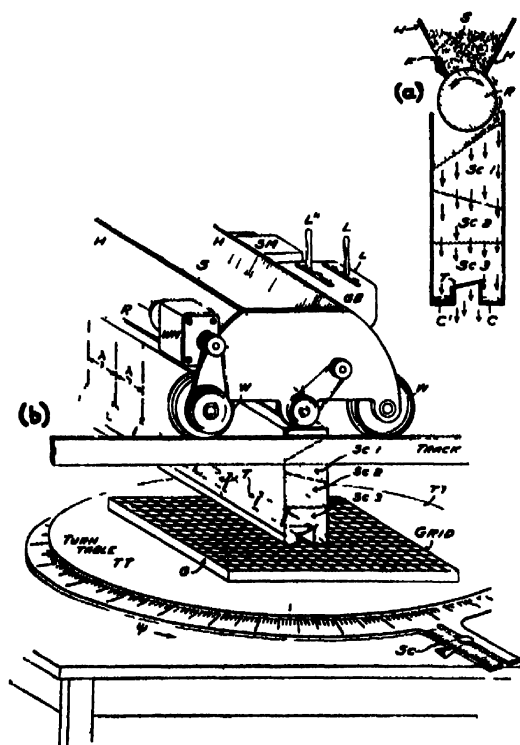


FIG. 4 A schematic drawing of the sand machine. (a) Diagram indicating action of the roller, screws, template and collector. (b) The sand machine. In this latter model, a variable speed motor is used instead of the gear box shown in Fig. 4.

¹ W. L. Bragg, *Zeits. f. Krist.* A70, 495 (1929).

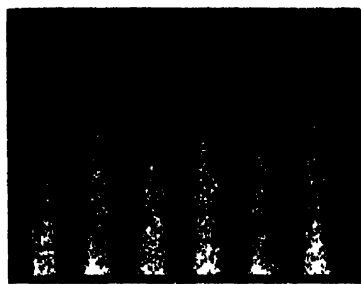


FIG. 5a. A single deposition of sand on felt.

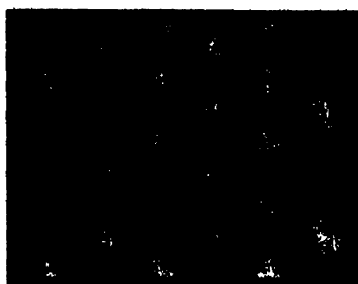


FIG. 5b. Two layers of sand; $\psi = 30^\circ$

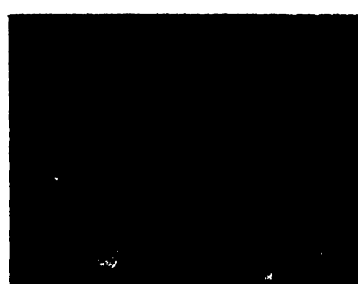


FIG. 5c. Three layers of sand (graphitic structure).

bottom of the truncated hopper. The rate that the sand pours from the hopper is determined by the speed of turning of the roller, this speed being governed by the settings of the levers L' , L'' , on the gearbox, GB , connected with the sand motor, SM . This method of obtaining an

TABLE I. Sample page of data for using the sand machine for projecting diopside.

Miller index hkl	d_{hkl} of template	Turntable angle ψ	Shift distance cm	F_{hkl} expressed in sand speed	
				Passes	Lever setting
002	13.8	0	6.9	4	2H
004	6.7	0	0	3	2H
402	8.5	321	4.25	{2 1	{2H 1H}
600	8.5	74	0	{2 1	{2H 3L}
404	5.4	342	0	2	2H
202	10.2	338	5.1	2	2H
602	8.5	71	4.25	{1 1	{2H 1H}
402	11.0	56	5.5	{1 1	{2H 1H}
404	6.7	31	0	2	1H
800	6.25	74	0	2	1H
408	3.95	343	1.97	2	1H
10, 02	4.4	304	2.2	{1 1	{2H 6L}
604	6.25	45	0	{1 1	{2H 3L}
10, 00	4.9	74	0	{1 1	{2H 3L}
802	5.2	307	2.6	{1 1	{1H 4L}
406	4.7	20	2.35	{1 1	{1H 2L}
202	13.8	32	0	1	2H
10, 02	5.2	82	2.6	1	2H
008	3.34	0	0	1	1H
206	4.7	10	2.85	1	1H
804	4.2	321	0	{1 1	{6L 3L}
802	6.7	79	3.85	{1 1	{5L 4L}
606	4.4	31	2.2	2	4L
006	4.4	0	2.2	2	4L
204	6.25	346	0	1	6L

even and adjustable flow of sand was found to be the only successful one of several methods tried. Figure 4a shows a cross section of the sand hopper and the roller. The sand is prevented from escaping from one side of the hopper by a strip of felt, F , while the other side of the hopper is slightly opened to permit the sand to ride through on the turning roller. The sand after escaping from the hopper falls through three 30-mesh screens so adjusted that after passage through the third screen the originally thin sheet of sand has been distributed over a band about two centimeters wide. The band of sand strikes a sinusoidal paper template, T . Since the template is tilted at an angle of about 40° , the sand which strikes it bounces back into the collector, C' . Some of the sand falls into C but the sand which goes through the openings in the template escapes in a sinusoidal distribution into the receiving grid, G . When the entire hopper arrangement travels along the track on the wheels, W , turned by the motor, WM , the sand is distributed in parallel bands analogous to the bands on Bragg's lantern slides.²

Since the grid is mounted on a graduated turntable, $T.T.$, any desired angle, ψ , can be obtained. The template has a sinusoidal wavelength, λ , proportional in centimeters to the spacing d_{hkl} in angstroms, and since the turntable can be shifted any desired fraction of d_{hkl} perpendicular to the direction of motion of the hopper, any desired value of phase α may be selected.

Figures 5a, b, and c show three successive depositions of sand on black felt at ψ values 120°

² W. H. Bragg and W. L. Bragg, *The Crystalline State* (Bell and Sons, Ltd., London, 1933), Vol. I, p. 231.

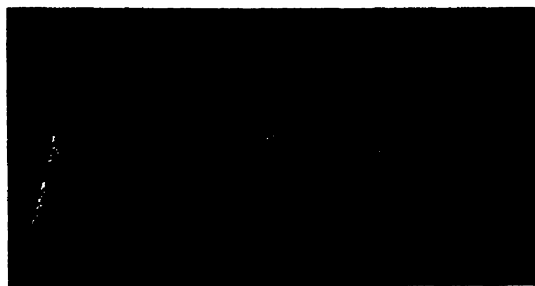


FIG. 6. Photograph showing the "hills" and "valleys" after the sand was dumped from the grid from a projection of diopside using Bragg's data (see reference 1).

apart to produce a structure very similar to that of graphite.

In using the machine it is convenient to have prepared a set of templates, each having a sinusoidal edge of the same amplitude of about one centimeter, but varying in wave-length by incremental steps between 0.8 centimeter to 20 centimeters. From the x-ray data the three quantities d_{hko} , ψ_{hko} , and F_{hko} and also the chosen value of α_{hko} are transformed to a table under the respective headings (a) template spacing, (b) turntable angle, (c) roller speed, and (d) shift distance. Table I shows a sample portion of such a tabulation of the data for the diopside structure. In this case the interplanar spacings d_{hko} were multiplied by a factor greater than the usual 1 centimeter per angstrom.

Illustrations of the Results from the Sand Machine

Figure 6 shows the sand after being poured out of the grid subsequent to making a projection of diopside.¹ It may be seen that there are "hills" corresponding to the positions of high electronic density where atoms are located and "valleys" between the hills. It is of academic interest to point out that, like the Bragg method, this method has no provisions for negative numbers since the sinusoidal distribution of sand for each traverse obeys Eq. (9) instead of Eq. (2).

$$S = [F_{hko}] + |F_{hko}| \cos [2\pi(hx + ky) - \alpha_{hko}], \quad (9)$$

and the final summation obeys Eq. (10) instead of Eq. (1).

$$\rho_{xy} = \sum_h \sum_k |F_{hko}| + \sum_h \sum_k |F_{hko}| \cos [2\pi(hx + ky) - \alpha_{hko}]. \quad (10)$$



FIG. 7. Resorcinol as projected on the sand machine using the data of Robertson and Ubbelohde (see reference 4).

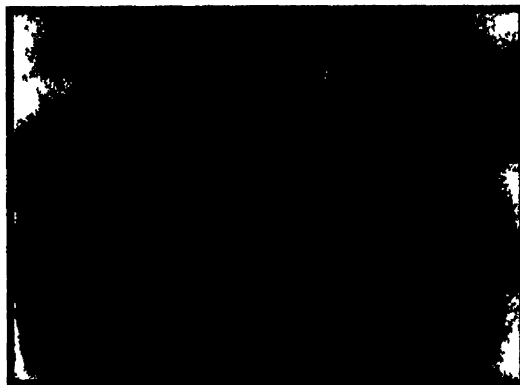


FIG. 8. Hexamethyl benzene projected into a grid having $\frac{1}{4}$ -inch spacings.

It was discovered that by illuminating the grid at an oblique angle from the sides while photographing it from directly above, a picture could be obtained which closely resembles the results obtained by Bragg¹ and Huggins.^{4,4} Figure 7 shows the resorcinol molecule photographed in this manner after being projected according to the dictates of the data of Robertson and Ubbelohde.⁵ Figures 8 and 9 show the hexamethyl benzene molecule using the data of Brockway and Robertson.⁶ These two pictures, differing in resolution, are presented to demonstrate the advantages of using a fine mesh grid, because of theoretical reasons discussed later in the paper.

¹ M. L. Huggins, J. Am. Chem. Soc. 63, 66 (1941).

⁴ M. L. Huggins, J. Chem. Phys. 12, 520 (1944).

⁵ J. M. Robertson and A. R. Ubbelohde, Proc. Roy. Soc. London A167, 122 (1938).

⁶ L. O. Brockway and J. M. Robertson, J. Chem. Soc. 1324 (1939).

THE ACCURACY OF THE METHOD

It is very difficult to derive a general law by which one may predict the accuracy to be expected from a Fourier summation in a crystal analysis on the basis of the known accuracies of the measured quantities used in the summation; nevertheless it should be understood that any machine which is developed for making the summations should be sufficiently accurate in its operation to make the fullest use of the accuracies of the x-ray data. Although it is known that in measuring the data from an x-ray film the accuracy of measuring the positions of the spots may be made extremely high (because these positions are so related to one another through the unit cell dimensions that statistics may be applied to their solution), the accuracies in measuring the intensities and ultimately the structure factor of the spots are not so easily handled. For example, if the crystal is not cylindrical in shape, the accuracy of the structure factors may be greatly dependent upon the accuracy of measuring the crystal dimensions. It is also observed that in ordinary microdensitometry the weakest diffraction spots are so badly distorted in value by the general background that no degree of accuracy can be claimed. Weak diffraction spots do not follow any general law in their location on the film but are very abundant in the regions of high Miller indices. Therefore, the best one can do in setting for himself a goal to be attained for a machine under construction is to make what may seem a reasonable guess. The goal chosen in this development is based on the assumption

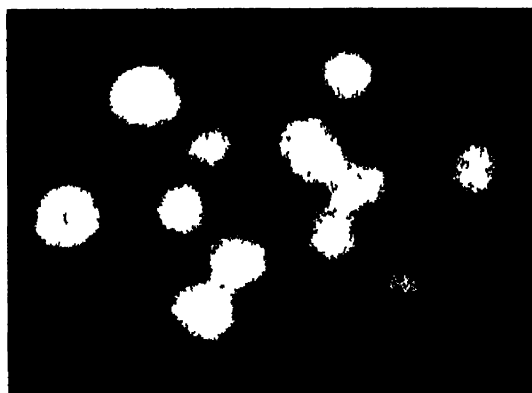


FIG. 9. Hexamethyl benzene projected into a grid having $\frac{1}{4}$ -inch spacings.

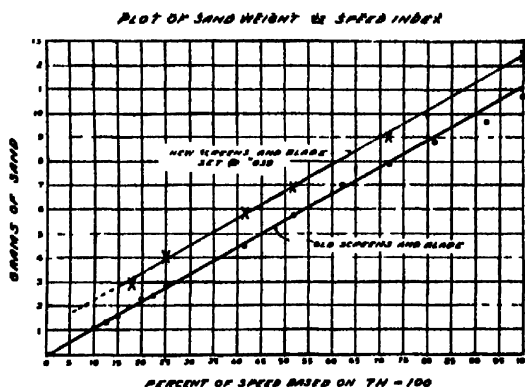


FIG. 10 Rate of sand flow plotted against the speed of rotation of the rod.

that the accuracy of the structure factors is 90 percent or greater for all spots corresponding to an interplanar spacing of one angstrom or more. To meet this goal, the amount of sand deposited by the machine in any single cavity in the grid should have an accuracy of 90 percent or more at the peak of a wave. The degree to which the various parts of the machine approach the goal are discussed separately.

The Rate of Sand Flow

The statement that the rate of sand flow is proportional to the rate of turning of the rod (R , Fig. 4) may be tested by experiment. In Fig. 10 the weight of sand discharged from the hopper in a set interval of time is plotted against the speed of rotation of the rod expressed as the percent of the fastest speed of rotation. The lower curve follows a straight line rather closely except at the faster speeds. This curve also passes through the origin corresponding (as one may expect) to zero flow at zero rate of turning. In the upper curve the knife blade above the rod was opened up to give a faster flow rate. This curve does not pass through the origin but indicates that there is a tendency for the sand to roll out independent of the peripheral motion of the rod. When the opening between the rod and the knife blade is too great, the slightest movement of the rod starts a small but continuous avalanche of sand. The second interesting characteristic of the upper curve is that its slope is very nearly the same as that of the lower curve. This suggests that the rotating rod has a

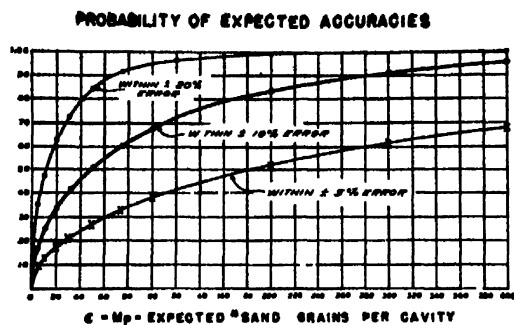


FIG. 11. A plot showing the probabilities of attaining inaccuracies within 5 percent, 10 percent, and 20 percent as a function of the number of grains of sand deposited in one grid cavity.

very definite depth of "bite" into the sand resting on it and when the opening is greater than this depth the rate of rotation is not a reliable measure of sand flow. In this work the opening of the knife blade has been kept below this upper limit. There is a lower limit for the size of the opening (at which point clogging occurs) which is easily avoided when sand grains of 100–140 mesh are used.

The Amount of Sand Required

If it were not for the ever present effects of statistical fluctuations in physical processes, it might be feasible to choose as the rate of flow of sand from the hopper for the lowest recorded structure factor that rate which would permit some small number (such as 10) grains of sand to fall into each cavity in the grid at the maxima of the waves and base all other rates on this standard. This is not possible. If the sand showers from the hopper through the screens and template into the grid in a perfectly statistical manner, there is a finite possibility that a cavity (into which 10 grains of sand were expected to have fallen) may be empty. According to probability theories, the percentage deviation from expectancy becomes smaller as the number of trials increases. It is required that we deposit into the grid enough sand to assure that any cavity in the grid at a wave crest will have a 9/10 probability of containing within 10 percent of the number of grains of sand that are expected. For this work we have used the normal law* and

* As pointed out at the Northampton meeting of the Crystallographic Society by D. Harker, the Poisson equation is more appropriate than the normal law for this particular work. The computations were not repeated, however, because the two laws give almost identical values in the working regions of the curves.

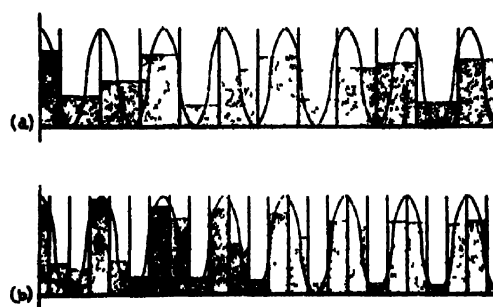


FIG. 12. (a) Distribution of sand in the cavities of the grid when $f=0.64$. (b) Distribution of sand in the cavities of the grid when $f=0.32$.

the tables in Fry's *Probability and Its Engineering Uses*. The central curve of Fig. 11 shows the probability of attaining an error of less than 10 percent plotted against the number of sand grains expected on the average to fall in each cavity in the grid. It may be seen that about 300 grains of sand are required to produce a 90 percent probability that the actual number of sand grains falling into any given cavity is within 10 percent of the number expected to fall in that cavity on the basis of a given rate of turning of the rod and a given rate of traverse of the hopper. The lower curve is added to show the tremendous number of sand grains required to attain an accuracy of 95 percent, and the upper curve shows the relatively few grains necessary to attain 80 percent.

The Size of the Cavities in the Grid

If the sand which drops through the template could be forced to "stick fast" to the exact position that it strikes the base of the grid, then partitions in the grid would be unnecessary. The fact that partitions in the grid force the sand to stay approximately in its initial position assures us that the use of grid sections only approximately enforces the desired effect. Since the accuracy is greater for smaller grid cavities, it is important to decide how small they should be.

When sand falls into a grid cavity, the vibration of the mechanism causes the sand to become leveled out although the sand may have de-

tion is more appropriate than the normal law for this particular work. The computations were not repeated, however, because the two laws give almost identical values in the working regions of the curves.

posited initially in a manner far from level. For this reason, the grid behaves as a sampling and averaging mechanism over an interval of area. For example, suppose that a curtain of sand were deposited across a grid according to the cosine equation

$$S = F_0 + F_0 \cos (2\pi/d_{hk0})P, \quad (11)$$

where P is the distance from the origin. Figure 12a shows a curve of this type. If a grid is used to sample and average such a curve, the height of sand in any cavity is

$$h_m = F_0 + \frac{F_0}{2\pi f} \sin 2\pi f(m + \frac{1}{2}) - \sin 2\pi f(m - \frac{1}{2}), \quad (12)$$

where f is the relative width, w , of the grid spacing expressed as a fraction of the wavelength d_{hk0} .

$$f = w/d_{hk0}, \quad (13)$$

and m is the serial number of each cavity when the cavities are numbered from left to right as shown in Fig. 12a. In Fig. 12a the heights of sand are shown as the dotted areas with grid separations so spaced that $f=0.64$. It can be seen that the height of the sand in the grid does not bear a very close relationship to the amount of sand actually deposited at specific points. Figure 12b shows the improvement obtained by using $f=0.32$. The reason that these specific values of f are used in the illustrations is that $f=0.64$ is the value resulting from the use of the quarter-inch grid for the 1 angstrom spacing and $f=0.32$ resulted from the one-eighth inch grid. A comparison of the hexamethyl benzene projections, Figs. 8 and 9, demonstrates the practical results obtained from small grid cavities. It may be stated that when f is less than 0.5 the largest error in representing a wave crest by a grid is

$$\text{percent } E = 100 \left(0.5 - \frac{\sin 2\pi f}{4\pi f} \right). \quad (14)$$

Using this equation, Fig. 12b should show only a 73 percent accuracy which may be verified by reading the accuracy directly off Fig. 12b at its worst peak at $m=18$. By actually solving for that value of f which permits an accuracy of 90 percent one finds that f must be about 0.18. This

means that the eighth-inch grid spacing is about twice as great as it should be for a scale of one angstrom per centimeter. Such fine grids, however, are more difficult to assemble, and the accuracy of 90 percent is obtained with an eighth-inch grid by enlarging the scale to one angstrom per two centimeters.

The Depth of the Grid

One may readily calculate the depth that the grid must be to assure that it will not run over before a summation is complete. Choosing a grid size of one-eighth inch square and knowing that 300 grains are the minimum number for the peak of the weakest wave, one may assume that a summation might require 200 wave additions with structure factor values whose ratios range from 1 to about 12 with a mean value of about six times the smallest measured value. If the sand has an average radius of about 0.01 cm and when loosely packed incorporates about 40 percent void space, then the total depth of sand in a cavity in which all 200 waves should contribute their maximum is about $1\frac{1}{2}$ inches.

The Required Number of Templates

The accuracy required in cutting the templates is also determined by accuracy of the x-ray data. If one wishes to have on hand a series of templates covering the range of wave-lengths from 0.8 cm to 20 cm so that he is prepared to project any structure whose interplanar spacings fall in this range, and if the accuracy of measurement of the F values is 90 percent, then there is no need to cut the templates more accurately than 93 percent. The choice of 93 percent or $(100-7)$ percent is based on the fact that the peak value

TABLE II. A proposed set of template wave-lengths.

Tem- plate No.	λ cm	Tem- plate No.	λ cm	Tem- plate No.	λ cm	Tem- plate No.	λ cm	Tem- plate No.	λ cm
1	0.80	11	1.56	21	4.52	31	8.84	41	17.34
2	0.86	12	1.72	22	4.83	32	9.46	42	18.56
3	0.90	13	1.84	23	5.16	33	10.12	43	19.86
4	0.96	14	1.97	24	5.52	34	10.80		
5	1.03	15	2.10	25	5.90	35	11.56		
6	1.12	16	3.24	26	6.31	36	12.37		
7	1.20	17	3.46	27	6.75	37	13.23		
8	1.28	18	3.70	28	7.22	38*	14.15		
9	1.37	19	3.96	29	7.72	39	15.14		
10	1.46	20	4.23	30	8.26	40	16.20		

TABLE III. Description of parts of the sand machine.

Sand	Ottawa, 100-200 mesh (obtained from American Grade Sand Company, Chicago, Illinois).	T	Templates, $4\frac{1}{2}$ cm over-all width, $30\frac{1}{2}$ " length, peak amplitude 3 cm, leaving $1\frac{1}{2}$ cm for holding (cut from heavy weight graph paper).
H	Sheet metal hopper, 7" wide at top, $4\frac{1}{2}$ " deep, $31\frac{1}{2}$ " long, sides tilt approximately 30° and converge at roller, R.	WM	Carriage motor drive, Bodine, 10 r.p.m., 30 in. oz., reversible, drive sprocket—16 teeth, driven sprocket—14 teeth.
R	Steel tubing, $\frac{1}{4}$ " wall, approximately .540" O.D., $31\frac{1}{2}$ " long.	Carriage assembly	Face plates, 12" wide, wheels $1\frac{1}{2}$ " diameter, $\frac{1}{2}$ " thick, flange $\frac{1}{2} \times \frac{1}{2}$ ", C.R. steel, 2 wheels on one shaft for drive, 2 wheels mounted on independent shafts, all wheels mounted on $\frac{1}{2}$ " roller bearings.
SM	Bodine Universal Motor, Thyatron controlled continuously between 15 and 200 r.p.m. (Gerald Heller Product, Baltimore, Maryland).	Track	$1\frac{1}{2} \times 1\frac{1}{2} \times \frac{1}{2}$ " iron T-bar stock, 56" long mounted $6\frac{1}{2}$ " from table on $1\frac{1}{2}$ " round brass posts with flanges attached for mounting to table.
K	Knife, $\frac{1}{2}$ " sq. C.R. steel bar, supported in bearings just above roller, R; one edge rotated to or from the roller to adjust opening to about .025".	Table	32" high, $61\frac{1}{2}$ " long, 43" wide, constructed of 4×4 " legs, $\frac{1}{2}$ " framing, and $\frac{1}{2}$ " plywood top surfaced with $\frac{1}{2}$ " Masonite.
Sc	Screens, No. 30 mesh, $2\frac{1}{2}$ " wide, stainless steel, fastened to $\frac{1}{8}$ " brass supporting bars which rest in bearings allowing tilt of screen for optimum sand distribution; 2 top screens tilt about 20° in same direction, bottom screen horizontal. All screens 3" apart, screen chamber about 7×3 ". Both sides of screen chamber are closed in with sheets of cellulose acetate to prevent sand escaping to outside.	T.T.	Turntable, $\frac{1}{2}$ " aluminum, 30" diameter, indexed at the edge throughout 360° , pivots on $\frac{1}{2}$ " pin, the latter supported by a $\frac{1}{2} \times 4 \times 40$ " slide strip which slides between fixed runners, also indexed on one edge in intervals of 1 mm over 15-cm distance.
	Note: Ends of hopper, H, and screen chamber are of $\frac{1}{2}$ " brass stock.	G	Grid, 20" square, $\frac{1}{2}$ " deep, constructed of 002" corrugated brass shim stock laid side by side with alternate partitions to form $\frac{1}{2}$ " cells supported by a $\frac{1}{2}$ " angle iron frame, mounted on 4 posts projecting from the turntable and high enough to provide clearance for a $\frac{1}{2}$ " plywood false bottom which can be dropped a uniform distance from the bottom of the grid as the grid fills before completion of run; thus making the effective depth of the grid increase from $\frac{1}{2}$ " to about $1\frac{1}{2}$ ".
C	Sand collector, $1\frac{1}{2}$ " deep, $4\frac{1}{2}$ " wide, $31\frac{1}{2}$ " long; contains template holder consisting of two $\frac{1}{2}$ " brass bars both mounted on edge to restrict a $\frac{1}{2}$ " opening along the center of collector, one bar 1.0" wide and the other $1\frac{1}{2}$ " wide. The difference in the two bars creates an angle, approximately 40° , across the two edges on which to rest the paper template.		

of a cosine curve is in error 10 percent at about 25° or about 7 percent of a cycle. Table II furnishes a proposed set of template wave-lengths, each wave-length being about 7 percent greater than the preceding one. One then selects, from the prepared set, that template which is the nearest in value to the wave-length desired.

In Table III the parts of the sand machine are given for convenience in duplicating this apparatus. The captions in this table correspond to those used in Fig. 4.

ACKNOWLEDGMENTS

The authors wish to thank the American Cyanamid Company for furnishing the facilities for carrying on this work and Dr. R. Bowling Barnes, director of the Physics Division of the Stamford Research Laboratories, for his enthusiasm and encouragement. We are also indebted to the Mining Division and to a number of machinists, draftsmen, and photographers of these Laboratories for their assistance and fruitful suggestions.

Radar Reflections from Long Conductors*

F. BLOCH

Stanford University, Palo Alto, California,

M. HAMERMESH

New York University, University Heights, New York, New York,

AND

M. PHILLIPS

Brooklyn College, Brooklyn, New York

(Received August 12, 1946)

The distribution of the backscattered radiation of a long straight conductor shows a sharp return lobe perpendicular to its orientation. In horizontal observation such a "rope" would therefore give an appreciable echo only if it were almost exactly vertical. In practice no such effect will be observed since, because of wind and other disturbances, the rope will never be straight but actually over its whole length will pass through regions of space where the phase of the incident wave will change many times through its full range 2π . The effect of these irregularities will be to make the sharp dependence of the cross section upon the orientation of the rope disappear, and to give a cross section proportional to the total length L of the rope. In Section I, the wave-length λ is assumed to be large compared to the transverse dimensions of the rope. In Section II, the results are extended to include the case where the wave-length λ of the incident radiation is not necessarily large but is comparable with or even small compared to the width d of a ribbon-shaped rope. It is found that the cross section of such a rope can be written generally in the form $\sigma = KA f(\pi d/\lambda)$, where K is a numerical constant depending upon the deviations of the rope from the vertical and where $A = Ld$ is the area of the ribbon. The length L is assumed to be much greater than λ . The function f is evaluated by the use of a method developed by Morse and Rubenstein.

DURING the past few years the design of radar reflectors has been studied extensively. These reflectors are used to produce strong echoes at the radar—as identifying or locating signals, or to confuse the enemy radar operator by presenting large numbers of false targets in addition to the true target. When such reflectors are to be used by aircraft, considerations of compactness, weight, and ease of handling are paramount in importance. This has, except for special applications, prevented the use of flat-sheet or spherical reflectors. Instead resonant dipoles have been used. If such dipoles are cut properly so as to resonate to the radar frequency, they will produce large intensities of reflection per unit weight. At longer wave-lengths, however, the resonance length becomes inconvenient and the dipoles tend to break. In this long wave-length region, long rolls of metal-

lized strip have been used. These long strips have been referred to as "ropes." The material is wound on spools in lengths of several hundred feet; after the spool is dropped from the aircraft, simple attachments (e.g., a cardboard attached to the free end of the rope) can be used to cause the roll to stream off the spool. In practice, ropes have usually been devised so that they are suspended approximately vertically during their fall.

In designing such reflectors one must know the value of the "radar cross section," σ , of the rope, i.e., the cross section for backscattering to the radar. Though in most cases the transverse dimensions of the rope (e.g., the width of the strip) are small compared to the wave-length, the dependence of the cross section on the transverse dimensions is of interest.

I

*This work reported in this paper was done at the Radio Research Laboratory of Harvard University under contract with the Office of Scientific Research and Development, National Defense Research Committee, Division 15.

We consider first a circular cylindrical rope, of radius a small compared to the wave-length λ ,

and of length l large compared to λ . We assume the rope to be straight and vertical and to be irradiated by a vertically polarized plane wave incident in the horizontal direction. The electric field, E , of this plane wave will be parallel to the axis of the rope and uniform along its length. The current induced in the rope will, except for end-effects which are unimportant since $l \gg \lambda$, also be uniform along the rope. If the material of the rope is perfectly conducting, the tangential component of the *total* electric field must vanish on the surface of the material. If I is the total current, then this condition is expressed by the equation

$$\frac{Ik}{2\pi ic} \int_0^{2\pi} d\varphi \int_{-\infty}^{+\infty} \frac{\exp [ik(z^2 + \rho^2)^{1/2}]}{(z^2 + \rho^2)^{1/2}} dz = -E, \quad (1)$$

where z is the coordinate along the rope and $k = 2\pi/\lambda$.

For $\rho \ll \lambda$ ($k\rho \ll 1$),

$$\int_{-\infty}^{+\infty} \frac{\exp [ik(z^2 + \rho^2)^{1/2}]}{(z^2 + \rho^2)^{1/2}} dz \approx 2 \ln (2/\gamma k \rho) + i\pi,$$

where $\gamma = 1.781$. From Fig. 1, $\rho = 2a \sin (\varphi/2)$, so that

$$E = -\frac{2Ik}{ic} [\ln (2/\gamma ka) + i\pi/2]. \quad (2)$$

The current I , uniform along the rope of length l , produces at a distance R , perpendicular to the rope and large compared to l , a field E' given by

$$E' = -\frac{k}{ic} \frac{e^{ikR}}{R} Il. \quad (3)$$

The cross section is defined as

$$\sigma = 4\pi R^2 |E'|^2 / |E|^2,$$

so that

$$\sigma = \frac{\pi l^2}{[\ln (\lambda/\pi \gamma a)]^2 + (\pi/2)^2}. \quad (4)$$

For arbitrary values of θ (the angle between the direction of incidence and the axis of the rope), and φ (the angle between the field E and

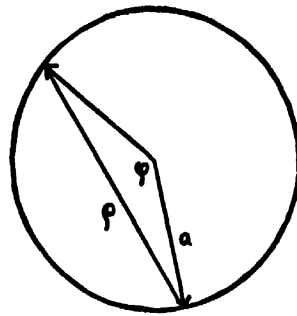


FIG. 1.

the plane of incidence), a similar derivation gives

$$\pi l^2 \sin^2 \theta \cdot \frac{1}{[\ln (\lambda/\pi \gamma a \sin \theta)]^2 + (\pi/2)^2} \cdot \left[\frac{\sin (2\pi l/\lambda \cos \theta)}{2\pi l/\lambda \cos \theta} \right]^2 \cdot \cos^4 \varphi. \quad (5)$$

These formulas have been obtained previously by L. J. Chu (unpublished).

If the rope is a thin metallic strip of width d , the results are changed only slightly, the quantity a is replaced by $d/4$ in Eqs. (4) and (5).

According to Eq. (5) the reflection pattern of a long straight rope has a sharp lobe covering an angular range $\lambda/2\pi l$ around $\theta = 90^\circ$. Thus for horizontal observation the rope will produce only a weak echo unless it is almost exactly vertical. In practice the rope will never be perfectly straight but will deviate from a vertical line by distances of several wave-lengths. These deviations will eliminate the strong angular dependence indicated by Eq. (5) and lead to a cross section more or less independent of the instantaneous shape and orientation of the rope, and proportional to its length l .

To derive such an "effective" cross section we shall assume that the rope consists of independently scattering sections, each of which is a full helical turn about the vertical axis. The angle α between the rope and the vertical is assumed to be small; the radius, r , of the helix is assumed to be large compared to λ . Let the vertical be the z direction and $x = x(z)$, $y = y(z)$ be the equation of the helix. For a wave incident along the x direction, the field at the rope will now be $Ee^{ikx(z)}$. The induced current will show this same phase variation along the rope provided

the phase varies only slightly within one wavelength along the rope. This condition is fulfilled if α is small. Then

$$I(z) = -\frac{icEe^{ikz(z)}}{2k[\ln(2/\gamma ka) + i\pi/2]}. \quad (6)$$

The field at a large distance R is

$$E' = \frac{k}{ic} \frac{e^{ikR}}{R} \int_0^h I(z) e^{ikz(z)} dz, \quad (7)$$

where h is the vertical projection of the helical turn. Thus

$$E' = -\frac{e^{ikR}}{R} \frac{E}{\ln(2/\gamma ka) + i\pi/2} \int_0^h e^{2ikz(z)} dz. \quad (8)$$

For a helix with angle α and radius r ,

$$x = r \cos(z/r \cot \alpha), \quad h = 2\pi r \cot \alpha.$$

Thus, if $\varphi = z/r \cot \alpha$,

$$E' = -\frac{Er \cot \alpha}{R \ln(2/\gamma ka) + i\pi/2} \int_0^{2\pi} e^{2ikr \cos \varphi} d\varphi. \quad (9)$$

For $kr \gg 1$, this becomes

$$E' \simeq -\frac{e^{ikR}}{2R} \frac{2\pi E \cot \alpha}{\ln(2/\gamma ka) + i\pi/2} \left(\frac{r}{\pi k}\right)^{1/2} \times \cos(2kr - \pi/4). \quad (10)$$

Thus the cross section will be

$$\sigma = 4\pi R^2 |E'|^2 / |E|^2 = \frac{4\pi^3 \cot^2 \alpha}{[\ln(2/\gamma ka)]^2 + (\pi/2)^2} \frac{r}{\pi k} \cos^2(2kr - \pi/4). \quad (11)$$

Since $kr \gg 1$ and we are interested only in average values, we may replace $\cos^2(2kr - \pi/4)$ by $\frac{1}{2}$. Also using $r = h/2\pi \cot \alpha$, $k = 2\pi/\lambda$, and $h = l \cos \alpha$, we obtain

$$\sigma = \frac{l\lambda}{2} \frac{1}{[\ln(\lambda/\gamma \pi a)]^2 + (\pi/2)^2} \frac{\cos^2 \alpha}{\sin \alpha}. \quad (12)$$

If now we consider a rope consisting of helical turns with varying l and α , we may assume that the scattering of the separate turns adds incoherently to produce the total cross section of

the rope. Thus we obtain the final formula

$$\sigma = K \frac{L\lambda/2}{[\ln(\lambda/\gamma \pi a)]^2 + (\pi/2)^2}, \quad (13)$$

where K is the average value of $\cos^2 \alpha / \sin \alpha$ (weighted in proportion to the length of the rope section), and L is the total length of the rope.

A quantitatively similar formula is obtained if, instead of the above procedure, one assumes the rope to consist of straight sections of varying length which contribute incoherently to the cross section. If one assumes that the inclinations to the vertical, α , are distributed uniformly between zero and a maximum value $\alpha_{\max}$, one obtains the same final expression (13), except that K is now $\pi/2 \sin \alpha_{\max}$. Obviously one cannot assign an exact numerical value to K without knowledge of the exact shape of the rope; qualitatively, at least, both procedures lead to the same result.

II

We can extend the calculation to the case where λ is comparable with or even small compared to the transverse dimensions of the rope. We shall consider the rope to be a flat strip of width d , and introduce the quantity $c = \pi d/\lambda$. ($c \ll 1$ is equivalent to $\lambda \gg d$.) Then Eq. (13) can be written as

$$\sigma = KA \frac{\pi}{2c} \frac{1}{[\ln(4/\gamma c)]^2 + (\pi/2)^2}, \quad (14)$$

where $A = Ld$ is the actual area of the ribbon. Equation (14) can be generalized for arbitrary λ by writing

$$\sigma = KA f(c). \quad (15)$$

Comparing with (14) we see that

$$f(c) = \frac{\pi}{2c} \frac{1}{[\ln(4/\gamma c)]^2 + (\pi/2)^2} \quad \text{for } c \ll 1. \quad (16)$$

A method for computing $f(c)$ for arguments c not necessarily small compared to one has been developed by Morse and Rubenstein.¹ Actually one finds in their paper an expression for the ratio of scattered intensity I' to incident inten-

¹ P. M. Morse and P. J. Rubenstein, Phys. Rev. **54**, 895 (1938).

sity I_0 at distances R , large compared to λ and d , but small compared to the length L of the rope. In this case the scattered intensity I' is of course proportional to $1/R$ whereas for distances large even compared to L it is proportional to $1/R^2$. Since the definition of a cross section can be meaningful only where the intensity follows a $1/R^2$ law, one must first convert their results.

Let I'/I_0 be the ratio of scattered to incident intensity for $R \ll L$, and I/I_0 the ratio for $R \gg L$. In either case the scattered intensity is proportional to the absolute square of the integral

$$H = \int_{-L/2}^{+L/2} \frac{\exp[ik(R^2 + z^2)^{1/2}]}{(R^2 + z^2)^{1/2}} dz. \quad (17)$$

For $kR \gg 1$, i.e., for distances large compared to the wave-length, this integral becomes

$$H = \left(\frac{2\pi i}{kR}\right)^{1/2} e^{ikR}, \quad \text{for } R \ll L. \quad (18)$$

If, on the other hand, R is so large that both $kR \gg 1$ and $kL^2/R \ll 1$, it becomes

$$H = (L/R) e^{ikR}. \quad (19)$$

The ratio of the absolute square of the two expressions (18) and (19) is

$$(2\pi/kR)(R^2/L^2) = 2\pi R/kL^2 = R\lambda/L^2.$$

Thus

$$I/I_0 = (L^2/R\lambda) I'/I_0. \quad (20)$$

Since I'/I_0 is proportional to $1/R$, we thus obtain the proper $1/R^2$ proportionality for I/I_0 . The factor L^2 which appears in (20) originates from the fact that this expression is derived for incidence and observation normal to a perfectly straight rope. Just as in Section I, even small deviations from straightness result in a scattered intensity and therefore a cross section σ which is proportional to the *first* power of L , or, as indicated in (15), to the area A covered by the rope. The result is that instead of L^2 in (20) there appears the factor $KL\lambda/2\pi$. Making this substitution and writing further $(I/I_0) = (\sigma/4\pi R^2)$,

we obtain from (20)

$$\sigma = 2KRL(I'/I_0), \quad (21)$$

which, with the expression of Morse and Rubenstein for I'/I_0 , leads to an expression of the form (15) for the cross section.

There is another feature of the problem that has to be taken into account before the results of Morse and Rubenstein can be applied. The scattered intensity will in general depend not only on the wave-length but also on the relative orientation of the plane of the ribbon with respect to the directions of incidence and observation. We are here interested only in return signals, where the direction of the scattered radiation is equal and opposite to that of incidence and is perpendicular to the longitudinal axis of the ribbon. In this case there remains only the dependence of the backscattered radiation on the azimuthal angle ψ by which the ribbon can rotate around its axis. As long as λ is large compared to d , i.e., $c \ll 1$, this dependence is negligible; even if the ribbon is viewed edgewise the strong diffraction of the waves around it makes it scatter almost as much as in front view. For $c \gg 1$, i.e., for λ small compared to the width of the ribbon, this dependence becomes, of course, very marked, since here one obtains a return signal only if the plane of the ribbon is practically perpendicular to the direction of incident and scattered radiation. In practice, the ribbon will always be twisted many times around its perpendicular axis. It is, therefore, the average value of I'/I_0 or of σ over the azimuthal angle ψ which is of practical interest. The function $f(c)$ which appears in (15) is thus always to be understood as this average value over ψ . $f(c)$ has been computed for various values of c , by use of the tables of Morse and Rubenstein.¹ (See Table I.)

The evaluation of $f(c)$ becomes rather troublesome for large values of the argument, since more and more terms in a certain double sum have to be evaluated. We therefore found it

¹ Cf. reference 1. The backscattering for electromagnetic waves with the electric vector parallel to the ribbon axis is derived from the first of Eqs. (12), p. 898, with the angle of incidence α equal to $\pi - \psi$ and the whole averaged over ψ . Similarly, the results for waves polarized so that the electric vector is perpendicular to the ribbon axis follow from Eq. (10), p. 897. The quantity which we call I'/I_0 is d/R times the expressions given in (10) and (12) of Morse and Rubenstein, averaged over ψ .

TABLE I.

c	.1	.2	.4	.6	.8	1.0	1.2	1.4
$f(c)$	1.085	.767	.576	.479	.426	.391	.369	.364

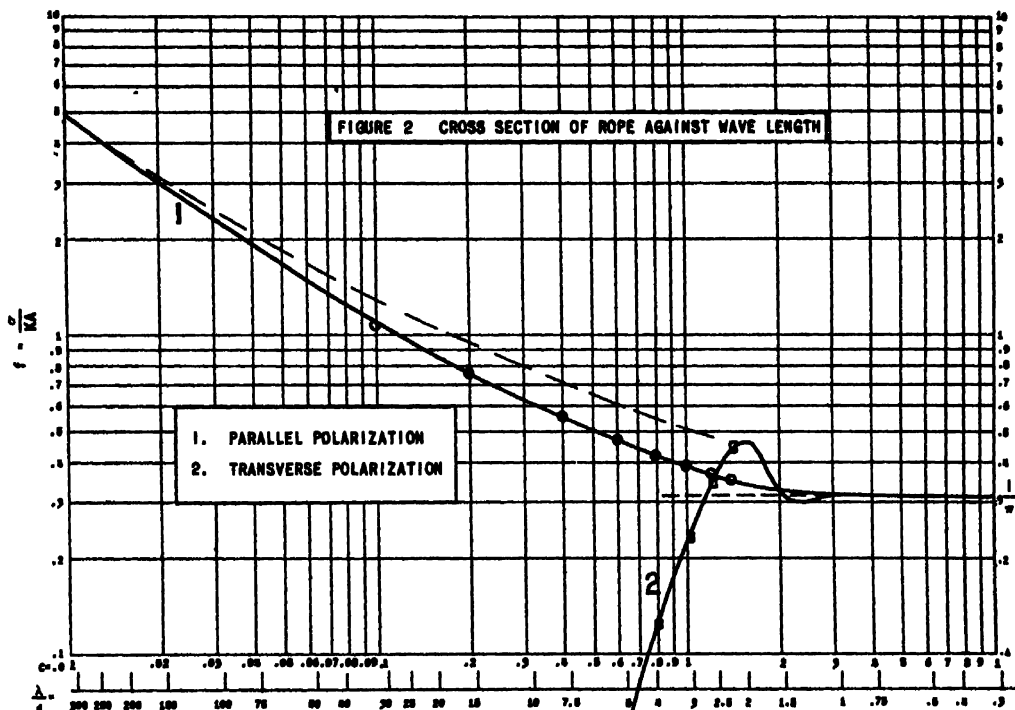


FIG. 2.

useful to derive the limiting expression of $f(c)$ for $c \gg 1$.

The method of obtaining this limiting value is essentially that used for the approximate treatment of diffraction phenomena. It consists of determining the current set-up by the external field in the ribbon at each point, neglecting the edge effects, i.e., treating it as though the ribbon were infinitely extended. From the expression for the current thus derived, one then determines the reradiated field by the usual radiation formulas. For a distance $R \ll L$, one thus obtains

$$\frac{I'}{I_0} = \frac{\sin^2 \psi \sin^2(kd \cos \psi)}{2\pi kR \cos^2 \psi}, \quad (22)$$

where ψ is the azimuthal angle of twist of the ribbon mentioned above, and where $\psi = \pi/2$ refers to an orientation of the plane of the ribbon perpendicular to the common direction of incidence and scattering. For $kd = 2\pi d/\lambda = 2c \gg 1$, one obtains an appreciable backscattering only in the neighborhood of the "specular" orientation $\psi = \pi/2$. The average value of (22) over ψ for $kd \gg 1$ gives

$$\langle (I'/I_0) \rangle_n = d/2\pi R. \quad (23)$$

Inserting this value for I'/I_0 in (21) we obtain

$$\sigma = \frac{K}{\pi} Ld = KA/\pi, \quad (24)$$

and, therefore, by comparison with (15) it is evident that

$$f(c) = 1/\pi \text{ for } c \gg 1. \quad (25)$$

While Eqs. (16) and (25) give the values of $f(c)$ in the two limiting cases $c \ll 1$ and $c \gg 1$, the intermediate region is covered by Table I, values computed from the tables of Morse and Rubenstein.

It is interesting to note that for $c = 0.1$ one obtains, from (16), $f(c) = 1.288$ instead of 1.085 and that the limiting value of $f(c)$ from (25) is 0.318, while for $c = 1.4$ it is actually 0.364. This means that Table I extends through a sufficient region so that both for smaller and larger values of c the much simpler limiting formulas (16) and (25) are good approximations. In curve 1 of Fig. 2 we have plotted $f(c)$ against c or λ/d , using the computed points of Table I and joining on to the dotted lines which represent the limiting expressions for small and large arguments.

TABLE II.

	.4	.6	.8	1.0	1.2	1.4
$f(c)$	.011	.075	.125	.237	.36	.46

So far we have assumed that the incident radiation is polarized parallel to the axis of the rope. As long as $\lambda \gg d$, any transverse component produces a negligible scattered intensity, of order $(d/\lambda)^4$ as compared to the parallel component. If, however, d becomes comparable to λ , the transverse component will also be scattered appreciably. In fact for $\lambda \ll d$, $f(c)$ will equal $1/\pi$ for either polarization. Table II gives the values of $f(c)$ for transverse polarization. These were obtained from the equations of Morse and Rubenstein in a manner similar to that for parallel polarization. These values are plotted in curve 2. For large arguments c this curve reaches the same constant $1/\pi$ as curve 1. While the value of $f(c)$ is negligible compared to that for parallel polarization below $c = \pi d/\lambda = \frac{1}{2}$, it

becomes equal for $c = 1.25$ and even exceeds it for $c = 1.4$. The numerical calculations have not been carried beyond that point and from there on the curve 2 is drawn somewhat schematically, keeping in mind only that it must approach the same value $1/\pi$ as curve 1. It seems plausible, however, that it should have a maximum in the neighborhood of $c = \pi d/\lambda = 1.5$, since this corresponds to a width approximately equal to $\lambda/2$. Since for transverse polarization the ribbon can be thought of as being composed of infinitely many infinitesimally narrow half-wave dipoles, the occurrence of such a maximum can be expected. Because of the strong coupling between these dipoles this maximum is quite broad and not very pronounced. For the same reason one can even expect additional maxima for larger arguments. Since they will be even less pronounced than the first maximum and have not been actually computed, they have been omitted in the drawing of curve 2. Even in the least reliable region between $c = 1.4$ and $c = 3$, we estimate the error to be no more than 20 percent.

Viscometric Investigation of Dimethylsiloxane Polymers*

ARTHUR J. BARRY

The Dow Chemical Company, Midland, Michigan

(Received July 29, 1946)

The intrinsic viscosities of an extended series of linear methyl polysiloxane fluids, having dimethylsiloxane (Me_2SiO) as the repeating unit, have been determined in toluene solution. These values have been correlated with their molecular weights, determined chemically and by osmotic pressure and light scattering measurements. The Staudinger equation was found applicable only for fluids of relatively low molecular weights, i.e., up to about 2500; for the higher polymers, the best correlation of intrinsic viscosity with number average molecular weight was found to be:

$$[\eta] = 2 \times 10^{-4} M_n^{0.66}.$$

This held reasonably well for molecular weights from 2500 to about 200,000. The bulk viscosities of the polysiloxane fluids were found to conform to the Flory relationship for melt viscosities; the expression:

$$\log \text{visc. (cs. at } 25^\circ\text{C)} = 1.00 + 0.0123 M_n^{0.4}$$

appears to be reasonably valid for molecular weights above 2500.

THE silicone fluids have recently become of considerable commercial and technical importance. They consist of a series of methylpolysiloxanes; details relating to their methods of preparation and description of properties are

now appearing in the literature. McGregor and Warrick¹ described a number of low and high polymer oils of cyclic and linear structures which were derived from the hydrolysis chiefly of dimethyldiethoxysilane. Intensive studies of the cyclic dimethylsiloxanes have been reported re-

* This paper, a contribution of the Dow Corning Corporation, was presented before the American Physical Society at the New York Meeting, January 26, 1946.

¹ R. R. McGregor and E. L. Warrick, U.S.P. 2,384,384 (1945).

cently by Hurd,³ Patnode and Wilcock,³ Hunter, Hyde, Warrick, and Fletcher,⁴ and Wilcock.⁵ The linear low polymers containing trimethylsiloxy groups as endblocking units are described in three of these papers^{3,4,5} while the high polymer members of this homologous series are described along with the low polymer species in a paper⁶ which has been submitted for publication.

The viscometric investigation of this important class of fluids, correlated with their molecular weight data, is of considerable practical as well as academic interest. The present paper is concerned principally with the linear members of the dimethylsiloxane polymers endblocked by ethoxyl or trimethylsiloxy groups.

TABLE I Methylpolysiloxane properties.

	Viscosity* (cs./25°C)	Intrinsic viscosity	Molecular weight	MW method	Calc'd MW**
Trimethylsiloxy endblocked species					
Hexamer	2.63	.0066	458	a, b	
Heptamer	3.24	.0077	532	a, b	
Octamer	3.88	.0088	606	a, b	
Nonamer	4.58	.0098	680	a, b	
Ethoxy endblocked species					
Heptamer	3.28	.0093	592	a, b, c	
Octamer	3.86	.0104	666	a, b, c	
Nonamer	4.50	.0119	740	a, b, c	
Decamer	5.17	.0124	814	a, b, c	
Ethoxy endblocked polymers					
	6.30	.0139	837	e	
	7.71	.0168	1,005	e	
	9.90	.0189	1,150	e	
	13.5	.0235	1,468	e	
	21.7	.0302	2,170	e	
	24.1	.0328	2,535	e	2,200
	37.4	.040	2,840	e	2,950
	48.0	.047	3,800	e	3,800
	76.8	.058	5,170	e	5,170
	82.0	.062	5,030	e	5,700
	102.0	.070	7,140	e	7,000
	156.0	.082	8,250	e	8,700
	290.0	.104	14,100	e	12,500
High polymer fluids					
	930.0	.159	25,300	s	23,500
	1,722.0	—	40,700	s	—
	1,698.0	.185	36,800	s	30,000
	1,960.0	.185	—	s	30,000
	3,060.0	.210	50,800	s	36,000
	10,600	—	70,800	p	—
	14,600	.310	—	s	65,000
	23,360	.340	69,900	s	75,000
	32,900	.374	—	s	87,000
	40,000	—	85,400	p	—
	65,000	.390	—	s	93,000
	125,000	.459	—	s	118,500
	310,000	.479	—	s	126,000
	746,000	.556	148,000	p	158,000

(a) Ultimate analysis.

(b) Boiling point elevation.

(c) Ethoxyl end-group analysis (reference 7).

(d) Osmotic pressure in methyl amyl ketone (reference 8).

(e) Light scattering in toluene (reference 9).

* Bulk viscosity of the silicone oil in centistokes at 25°C.

** Calculated from $[\eta] = 2.00 \times 10^{-4} M^{0.66}$.

³ C. B. Hurd, J. Am. Chem. Soc. 68, 364 (1946).

⁴ W. Patnode and D. F. Wilcock, J. Am. Chem. Soc. 68, 358 (1946).

⁵ M. J. Hunter, J. F. Hyde, E. L. Warrick, and H. J. Fletcher, J. Am. Chem. Soc. 68, 667 (1946).

⁶ D. F. Wilcock, J. Am. Chem. Soc. 68, 691 (1946).

⁷ M. J. Hunter, E. L. Warrick, J. F. Hyde, and C. C. Currie, paper submitted for publication.

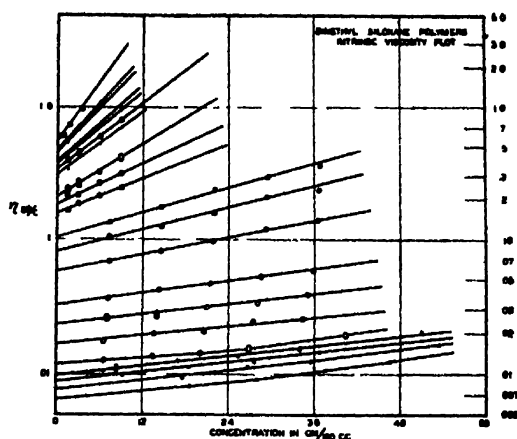
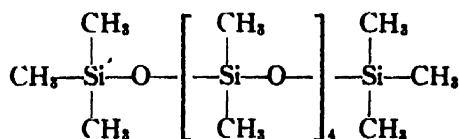


FIG. 1. Determination of intrinsic viscosities of methylpolysiloxane fluids.

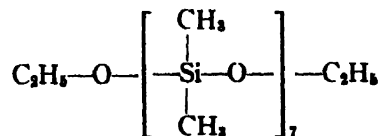
MATERIALS

Fluids of a wide range of viscosities and chain lengths were investigated as grouped in Table I. The first group is comprised of distilled species of trimethylsiloxy endblocked polymers of low molecular weight determined by ultimate analysis as well as by boiling point elevation. The first member listed has the structure:



Tetradecamethylhexasiloxane

The second group is comprised of distilled species of ethoxy endblocked polymers whose molecular sizes were established as above and confirmed by endgroup analysis. The first member listed has the structure:



Diethoxytetradecamethylheptasiloxane

Each fluid in the third group consists of a mixture of intermediate polymers of the preceding type. They were prepared by partial hydrolysis of diethoxydimethylsilane followed by aspiration at elevated temperatures to remove the distillable low polymer.⁷ The number average molecular weights were estimated from their

⁷ Unpublished work by M. J. Hunter and H. J. Fletcher.

ethoxyl analyses determined by a modified Zeisel method.⁸

The high polymer fluids were mixtures and were of molecular weights too great for sensitive determination by chemical methods of analysis. A number of them were characterized by average molecular weights estimated by osmotic pressure⁹ and light scattering¹⁰ methods using solutions in heptanone-2 and toluene, respectively.

EXPERIMENTAL

For this study, toluene solutions of the above samples were employed in concentrations up to 60 g/100 cc for the low polymer species, 37 g/100 cc for the intermediate polymers, and 9 g/100 cc for the high polymer fluids. The relative viscosities were determined at $20^\circ \pm 0.05^\circ\text{C}$ using Ostwald tubes of 1.8–2.3 cc effluent volume and 10-cm capillary length. The discharge times for solvent were generally in the range of 0.8 to 1.4 minutes though tubes of 3.8 and 12.5 minute solvent efflux times were frequently used for checks. A kinetic energy error of about 2 percent was calculated for the intrinsic viscosities so determined; the correction was not applied.

The specific viscosity was determined conventionally:

$$\eta_{sp} = \eta_r - 1 = \left(\frac{t \cdot d}{t_s \cdot d_s} - 1 \right)$$

where η_{sp} and η_r = specific and relative viscosities respectively, t = solution time, t_s = solvent time, d = density of solution, and d_s = density of solvent. Expressing the concentration, c , in g/100 cc, the values of η_{sp}/c were calculated and plotted on logarithmic scale against the concentrations as shown in Fig. 1. The intrinsic viscosity, $[\eta]$ the limiting value of η_{sp}/c as the concentration approaches zero, was read off the graph as the intercept upon extrapolation of each curve. The intrinsic viscosities are entered in Table I.

DISCUSSION

With the exception of the very low polymers, the curves for which show some curvature in the semi-logarithmic plot, the points representing η_{sp}/c as a function of concentration fall in straight

lines, permitting very good extrapolations and conforming to the Martin¹¹ relationship:

$$\log (\eta_{sp}/c) = \log [\eta] + k'[\eta]c.$$

While Martin reports the value of k' to be unique for a polymer-homologous series in a given solvent, this does not prove true in the present work. Here the value of k' varies as an inverse function of the size of the molecule. To illustrate, a few of the k' values are: 0.315, 0.222, 0.148, 0.135, 0.124 for the 7.7, 24, 290, 930, and 23,300 cs. fluids, respectively. They seem to approach 0.11 as a limit. The Huggins equation¹² does not yield a value of k' for any one polymer which is independent of solution concentration. If a limiting value of k' be determined for each polymer by extrapolating to zero concentration the plot of k' vs. concentration using the Huggins equation, it is found that the limiting k'_0 -values diminish through the polymer series in the same manner as the Martin k' -value. Thus for the 24, 290, 930, and 23,300 cs. fluids, the k'_0 -values were found to be 0.46, 0.33, 0.31, 0.29, respectively, approaching about 0.25 as an apparent limit for the series.

It is difficult to ascribe k' values to the lower polymer species of the first two groups, because of some curvature of their plots in Fig. 1. Nevertheless, some preliminary observations on the comparison of the plots of these open chain low polymers with those of a few corresponding cyclic polymers are of significance in the attempt to correlate k' values with structure. It was noted that the cyclic pentamer, hexamer, and octamer, $(\text{Me}_2\text{SiO})_n$, showed lower intrinsic viscosities than their linear counterparts while their (η_{sp}/c) vs. concentration slopes, and consequently their k' values, were greater. These data suggest that the compact rather fixed structure of the cyclic siloxanes contribute relatively little to flow resistance of dilute solutions yet have a pronounced effect in higher concentrations due to greater molecular interaction than is manifest in the open chain polymers. A greater molecular interaction is further reflected in the higher bulk viscosities of the cyclic polymers and in the

⁸ E. P. Sarnel and J. A. McHard, Ind. Eng. Chem. Anal. Ed. 14, 750 (1942).

⁹ R. M. Fuoss and D. J. Mead, J. Phys. Chem. 47, 59 (1943).

¹⁰ P. Debye, J. App. Phys. 15, 338 (1944).

¹¹ A. F. Martin, paper presented at the Memphis Meeting of the American Chemical Society, April 1942, "Cellulose and cellulose derivatives," edited by Emil Ott (Interscience Publishers, Inc., New York, 1943), p. 966.

¹² M. L. Huggins, J. Am. Chem. Soc. 64, 2716 (1942).

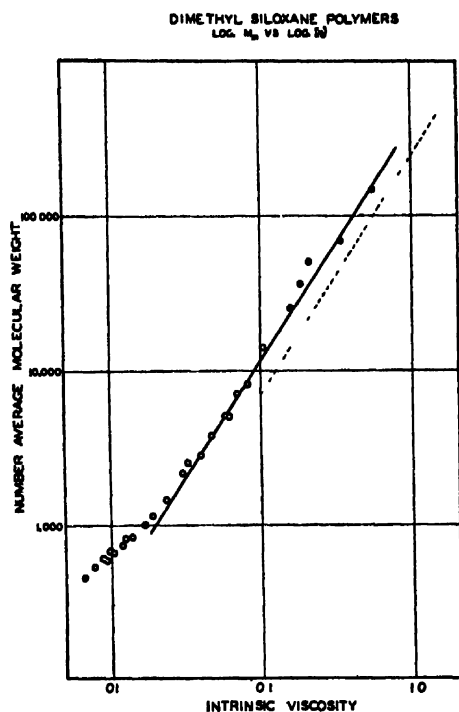


FIG. 2. Logarithmic plot of intrinsic viscosities of methylpolysiloxanes vs. molecular weights determined by: ○ chemical analysis; ⊕ light scattering; ● osmotic pressure. The dotted line represents Flory's data on polyisobutylene polymers (reference 14).

greater change of their viscosities with temperature change.

The first twelve polymers listed in Table I show exceedingly low intrinsic viscosities, which interestingly fall below the 0.025 limit of the Einstein equation when the concentrations are corrected to its fraction volume basis:

$\eta_r = 1 + 2.5 v/V$ or $\eta_{sp} = 2.5 v/V = 0.025 v/100 \text{ cc}$, where v = volume of solute and V = volume of solvent. Similar failure of the Einstein equation has been indicated previously by Meyer and Van der Wijk¹³ who reported on the viscosities of a series of paraffin hydrocarbons from $C_{17}H_{36}$ to $C_{34}H_{70}$ in CCl_4 solution.

The intrinsic viscosity data for the ethoxy endblocked species and the intermediate polymers up to a molecular weight of 1500 fit the simple form of the Staudinger equation $[\eta] = K \cdot M$ where $K = 1.62 \times 10^{-3}$. The intrinsic viscosities of the trimethylsiloxy endblocked species are somewhat lower and $K = 1.43 \times 10^{-3}$. To

¹³ K. H. Meyer and A. Van der Wijk, *Helv. Chim. Acta*, **18**, 1067 (1935); *Kolloid Zeits.* **76**, 278 (1936).

compensate for the fact that the chain of a polymer of the ethoxy endblocked type is longer, by two C_2H_5O -groups, than the corresponding member of the other series, it is interesting to correlate the intrinsic viscosity with the chain length, L , in atom units. If this is done, totaling all the C, O, and Si atoms in the chain, the relation becomes $[\eta] = K_1 \cdot L$ where $K_1 = 5 \times 10^{-4}$. This holds reasonably well for either type of polymer up to 80 atoms in length ($M = 2840$).

For the polymer mixtures of higher molecular weights, it was necessary to supplant the simple Staudinger equation by an exponential form. For such an evaluation, the intrinsic viscosities were plotted against the molecular weights using double logarithmic coordinates as shown in Fig. 2. The data fall on a reasonably straight line from which may be drawn the relationship:

$$[\eta] = K \cdot M^a,$$

where $[\eta]$ is the intrinsic viscosity, M is the number average molecular weight, K is a constant $= 2.00 \times 10^{-4}$, and a is an exponent $= 0.66$.

It is recognized that the molecular weights determined by light scattering are weight average values whereas all the other points are number average. Since the high viscosity fluids were commercial samples from which the lower polymers were removed by distillation at reduced pressure, the molecular weight distribution of the several species contributing to the average is upset from a normal probability function. Hence the conventional factor of $\frac{1}{2}$ for deriving number average molecular weights from weight average data is not applicable to the values determined by light scattering. However, there should be some negative correction in applying these molecular weight data to the graph which is otherwise drawn to number average molecular weights. With this qualification, the above relationship is quite reliable and fair practical agreement is shown between the reported molecular weights and the values estimated from the equation, as entered in the last column of Table I. A program involving fractionation of a high polymer mixture to provide specimens of narrow molecular weight ranges for similar study is projected.

It is interesting to view the analogy in structure between the polydimethylsiloxanes and the polyisobutylenes:

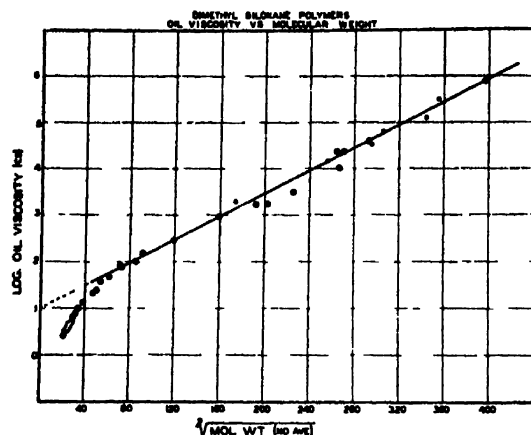
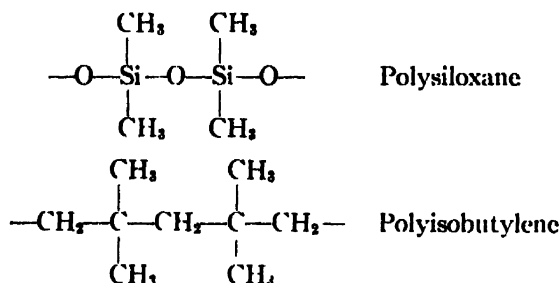


FIG. 3. Flory relationship plot for methylpolysiloxane fluids. Molecular weights determined by: ○ chemical analysis; ⊕ light scattering; ● osmotic pressure; ○ calculation from intrinsic viscosity equation.



Flory¹⁴ investigated a series of polyisobutylene fractions. Determining the viscosities in diisobutylene, he found the following equation to be valid for molecular weights above 5000:

$$[\eta] = K \cdot M^a,$$

where $K = 3.60 \times 10^{-4}$ and $a = 0.64$. This expression is represented in Fig. 2 by the broken line displaced toward higher intrinsic viscosities. The comparable slopes (a^{-1}) suggest a similar degree of chain kinking or spiral configuration for these structural analogs. This kinked structure may be reflected as one factor in the relative insensitivity to temperature of the viscosities, i.e., low viscosity/temperature slopes, observed for the polysiloxane fluids. Increase in temperature may tend to open out the kinked structure, invoking greater molecular interaction, and so compensate to some extent the normal decrease of viscosity with temperature rise.

When the logarithms of the oil viscosities (in centistokes at 25°C) were plotted against the

P. J. Flory, J. Am. Chem. Soc. 65, 372 (1943).

square roots of the molecular weights, as illustrated in Fig. 3, it was found that the Flory relationship¹⁵ for melt viscosities was applicable. Over most of the range, a straight line was obtained, the equation for which was found to be:

$$\log \eta_{(cs./25^\circ C)} = a + cM_n^{1/2}$$

where

$$a = 1.00, \quad c = 0.0123.$$

The expression appears to be reasonably valid for molecular weights above 2500. It should be noted that number average molecular weights were considered in the above formula. That a straight line plot was so obtained indicates that the several fluids are characterized by relatively narrow ranges of molecular weight or by comparable types of molecular weight distribution.

ACKNOWLEDGMENT

The author is indebted to Dr. M. J. Hunter, Mr. C. C. Currie, and Mr. H. J. Fletcher for providing the fluids studied. The author also wishes to thank Mr. R. F. Lind for osmotic pressure determinations and Mr. R. S. Spencer for light scattering measurements of molecular weight on several of the samples.

SUMMARY

The intrinsic viscosities of an extended series of linear methyl polysiloxane fluids, having dimethylsiloxane (Me_2SiO) as the repeating unit, have been determined in toluene solution. These values have been correlated with their molecular weights, determined chemically and by osmotic pressure and light scattering measurements. The Staudinger equation was found applicable only for fluids of relatively low molecular weights, i.e., up to about 2500; for the higher polymers, the best correlation of intrinsic viscosity with number average molecular weight was found to be:

$$[\eta] = 2 \times 10^{-4} M_n^{0.66}.$$

This held reasonably well for molecular weights from 2500 to about 200,000. The bulk viscosities of the polysiloxane fluids were found to conform to the Flory relationship for melt viscosities; the expression:

$$\log \text{visc. (cs. at } 25^\circ\text{C)} = 1.00 + 0.0123 M_n^{1/2}$$

appears to be reasonably valid for molecular weights above 2500.

¹⁵ P. J. Flory, J. Am. Chem. Soc. 62, 1057 (1940).

Theory of Small Signal Bunching in a Parallel Electron Beam of Rectangular Cross Section

EUGENE FEENBERG* AND DAVID FELDMAN**

Washington Square College, New York University, New York, New York

(Received August 8, 1946)

The phenomenon of bunching which takes place in a velocity modulated electron beam can be described in purely kinematical terms if space charge effects are negligible. However, the non-uniform distribution of charge in the bunched beam gives rise to a field which opposes the bunching process so that the kinematic solution may be validly applied only for a limited length of drift space. An accurate solution of the bunching process requires the integration of the dynamical and field equations. These reduce to a linear homogeneous system under the assumption of "small signal" conditions. The device of a high frequency "surface charge" is employed in formulating the boundary conditions at the surface of the beam. In the non-relativistic approximation the dynamical and field equations yield solutions which are classified as non-solenoidal or solenoidal accordingly as the motion produces or does not produce a high frequency charge density within the beam. There exists a non-solenoidal

type of motion which generates no field outside the beam. The actual physical problem is solved by taking a suitable linear combination of both solutions. Under conditions met with in practice, a large part of the actual solution may be of the solenoidal type. The high frequency component of beam current appears as a sum of a volume current and a "surface current," the latter term arising from the longitudinal motion of the high frequency "surface charge." Again, under conditions met with in practice, a large part of the high frequency component of beam current may occur in the form of "surface current." The theory contains three debunching wave numbers: one identical with that which occurs in Webster's debunching theory for a beam of unlimited cross section, and two which depend upon the transverse dimensions of the beam and drift tube, the current density, frequency, and beam velocity.

1. INTRODUCTION

AN adequate treatment of the bunching process in a velocity modulation tube must take account of the debunching action arising from the non-uniform distribution of charge in the bunched beam. Space charge effects become especially important in small signal amplifiers where the transit time in the drift space is made large in order to secure the maximum possible voltage gain. A satisfactory discussion of the bunching phenomenon therefore cannot be based on a purely kinematic treatment, but requires the integration of the dynamical and field equations which relate the motion of the electrons to the debunching field subject to suitable boundary conditions.¹ The method which is used here

follows very closely that employed by one of the authors for the treatment of small signal bunching in a parallel beam of circular cross section.²

In this paper we are concerned with a parallel electron beam of rectangular cross section centered in a rectangular drift tube. In order to simplify the analysis, we take the width of the beam to be infinite; this has the advantage of making the problem two-dimensional. The results that are derived on the basis of this assumption will hence be applicable in practice to a ribbon electron beam whose width is much greater than its depth. We make the following additional approximations:

- (1) Beam velocity/velocity of light $\ll 1$. (1)

This has the consequence that the equations of motion may be written in non-relativistic form; it also permits a simple description of the electromagnetic field generated by the bunching process.

¹ We make repeated reference to a volume on the theory and practice of klystron design and operation now in process of publication under the editorship of E. Ginzton. The authors of the volume are present and former members of the Tube Laboratory of the Sperry Gyroscope Company, Garden City, New York.

* Now at Washington University, St. Louis, Missouri.

** Now at Harvard University, Cambridge, Massachusetts.

¹ A method of solution in terms of an infinite set of characteristic solutions which must be combined in the correct proportions to obtain a solution of the actual physical problem has been given by W. C. Hahn, *Gen. Elec. Rev.* 42, 258 (1939); W. C. Hahn and G. F. Metcalf, *Proc. I.R.E.* 27, 106 (1939); S. Ramo, *Proc. I.R.E.* 27, 757 (1939). This procedure is especially well adapted to the case of an infinitely strong longitudinal magnetic field which completely suppresses all transverse motion. The opposite extreme of no magnetic field is the only condition considered in this paper.

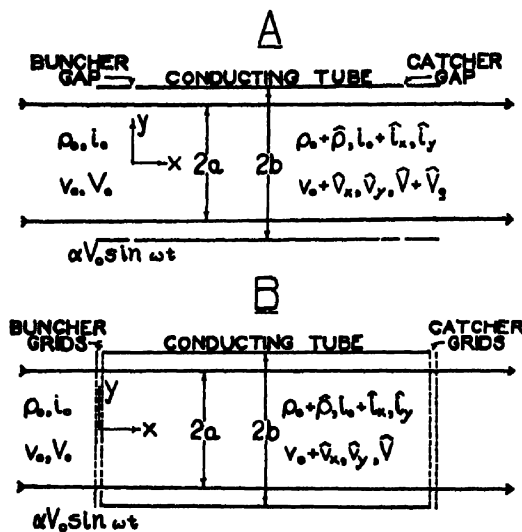


FIG. 1. A. Parallel ribbon electron beam centered between plane parallel conducting sheets. Gaps in the drift tube provide coupling to the input and output resonators. B. Parallel ribbon electron beam centered between plane parallel conducting sheets. Beam traverses grid sections which form part of the input and output resonators.

(2) Buncher voltage/beam voltage $\ll 1$ (2a)

and

High frequency component of beam current/total beam current $\ll 1$. (2b)

We are thus effectively restricting the discussion to the case of "small signals." Under these circumstances, the equations of motion and the field equations reduce to a linear system.

(3) We assume finally that there is an accumulation of positive ions in sufficient numbers to neutralize the time average field due to the bunched beam. In other words, in the absence of hf (high frequency) fields, there is no field at all acting on the unbunched beam in the drift space; in the presence of hf fields, the debunching action arises from the non-uniform distribution of charge in the bunched beam.

The type of coupling between the beam and the modulating hf field will influence the choice of boundary conditions and hence the results obtained to a large degree. We shall consider two types of coupling. In the first case (illustrated in Fig. 1A), coupling is achieved through gaps in the drift tube; each gap forms part of a resonant circuit or cavity resonator such that an electromagnetic field in the resonant circuit or cavity produces a field within the drift tube in the

neighborhood of the gap. In the second case (illustrated in Fig. 1B), the drift tube is closed at either end by a pair of closely spaced plane grids (each such pair forming the capacity section of a re-entrant cavity resonator) which serve to couple the beam to the buncher and catcher fields, respectively.

2. LIST OF SYMBOLS

All electrical quantities are stated in m.k.s. units. The subscript 0 and boldface type designate steady-state and the hf components, respectively; the additional subscripts x and y distinguish between longitudinal and transverse components. The following symbols are of importance in the calculations:

V_0	The potential in the charge free drift space resulting from an hf voltage across the buncher gap. (Note that $V_0 = 0$ for a tube with grids.)
V	The additional potential resulting from the hf fluctuations in charge density. (V is the total hf potential in the drift space for a tube with grids.)
v	Velocity.
ρ	Charge density.
$i_x = (\rho_0 + \hat{\rho})(v_0 + \hat{v}_x) - \rho_0 v_0$	Longitudinal hf current density.
$i_y \approx \rho_0 \hat{v}_y$	Transverse hf current density.
D	Transit angle in radians across the buncher grid space or gap opening.
$\mu = 2/D \cdot \sin D/2$	Beam coupling coefficient.
$\mu_x = \mu \cosh(\eta y/a) / \cosh(\eta b/a)$	Beam coupling coefficients in a gap system.
$\mu_y = \mu \sinh(\eta y/a) / \cosh(\eta b/a)$	
λ'	"Beam wave-length" or the average distance traveled by an electron in one cycle.
$k' = 2\pi/\lambda'$	"Beam wave number."
$\eta = k'a, \eta' = k'b$	
$G_0 = 2ai_0/V_0$	Beam admittance per unit width of the beam.
$\beta_0 = v_0/c$	Ratio of beam velocity to the velocity of light.
$h = (\rho_0/2\epsilon_0 V_0)^{1/2}$	Debunching wave number. To compute h we use the equivalent expression where G_0 is in mhos per meter.
$h\alpha = (30\pi a G_0/\beta_0)^{1/2}$	
α	Ratio of hf voltage to beam voltage.
τ	Phase at which an electron enters the drift space ($x=0$).
$\langle f \rangle$	Average value of the function f over the cross section of the beam.

Throughout most of the analysis, it will prove to be useful to express the potential, velocity, current density, and charge density in terms of a basic set of units in order to effect a consider-

able simplification in the appearance of many of the equations. The set used consists of $2V_0$, v_0 , i_0 , and ρ_0 for the potential, velocity, current density, and charge density, respectively. To transform an expression involving reduced units into ordinary units, we must then make the following substitutions:

$$\begin{aligned} V &\rightarrow V/2V_0, \\ v &\rightarrow v/v_0, \\ i &\rightarrow i/i_0, \\ \rho &\rightarrow \rho/\rho_0. \end{aligned}$$

Unless otherwise stated, reduced units are employed.

3. FIELD EQUATIONS AND EQUATIONS OF MOTION

In view of the assumption contained in Eq. (1), the magnetic field of the beam may be neglected in computing the force acting on an electron in the beam; furthermore, the electric field may be derived from a quasistatic potential function $V(x, y, t)$. The net error in the expression for the force due to these two simplifications consists in the neglect of a term of order β_0^2 relative to the leading term which is retained.³ The resulting field equations are quite analogous to those which hold for an electrostatic field, *viz.*,

$$\mathbf{E} = -\nabla V \quad \text{and} \quad \nabla \cdot (\epsilon_0 \mathbf{E}) = \rho,$$

so that

$$\Delta V = -\rho/\epsilon_0$$

in ordinary units, and in reduced units

$$\Delta V = -k^2 \phi. \quad (3)$$

Conservation of charge may be expressed by the equation

$$\nabla \cdot \mathbf{i} + \partial \rho / \partial t = 0$$

in ordinary units, which in terms of our reduced units becomes

$$v_0 \nabla \cdot \mathbf{i} + \partial \phi / \partial t = 0. \quad (4)$$

We have finally the equations of motion

$$m \dot{\mathbf{v}}_x = -e \partial V / \partial x, \quad m \dot{\mathbf{v}}_y = -e \partial V / \partial y$$

in ordinary units, which become, on substituting

³ See, for example, the discussion of the forces acting on and due to electrons in uniform motion given in Jeans, *The Mathematical Theory of Electricity and Magnetism* (Cambridge University Press, Cambridge, 1945), fifth edition, p. 560.

reduced units,

$$\dot{\mathbf{v}}_x = -v_0 \partial V / \partial x, \quad \dot{\mathbf{v}}_y = -v_0 \partial V / \partial y. \quad (5)$$

Equations (3)–(5) constitute the theoretical system for our problem.

Throughout the remainder of this report, we make use of the fact that in practice $h \ll k'$; we assume this inequality in the form $ha \ll \eta$. In seeking solutions of Eqs. (3)–(5), it therefore proves convenient to employ a perturbation procedure with the quantity ha/η as expansion parameter. A first-order perturbation calculation is all that is desired so that we shall consistently neglect terms of order $(ha/\eta)^2$ with respect to the leading term in all of our equations. On the other hand, it should be noted that it is not necessarily true that hx is a small quantity, whence all powers of hx must be kept in the expressions for hf quantities.

An analogous situation occurs in Webster's kinematic bunching theory⁴ which employs the quantity α as expansion parameter in a first order perturbation calculation. Here, terms of order α^2 with respect to the leading term may be neglected in all expressions. However, the quantity $\alpha \cdot k'x$ which is simply the product of α by the transit angle in radians in the drift space is not necessarily small, so that all powers of $\alpha \cdot k'x$ must be retained.

4. BOUNDARY CONDITIONS

We next impose boundary conditions on the various quantities appearing in (3)–(5), *viz.*, on the potential function, the entering hf components of the longitudinal and transverse velocities, and the entering hf component of beam current density (or charge density). The specific nature of the boundary conditions will be determined by the type of coupling between the beam and the hf modulating field.

The gap system has been treated in detail.⁵ It has been shown that the velocity modulation produced by an hf gap may be expressed in terms of the beam coupling coefficients μ_x and μ_y as though produced instantaneously at the plane $x=0$. Actually the gap field is confined to a region about the plane $x=0$, attenuating exponentially in either direction. An explicit consideration of V_x is unnecessary however since the

⁴ D. L. Webster, J. App. Phys. 10, 501 (1939).

beam coupling coefficients μ_x and μ_y essentially express the over-all effect of V_g .

For a gap system, we therefore have the following boundary conditions:

$$\left. \begin{aligned} V(x, b, t) = V(0, y, t) = 0 \\ v_x = \frac{1}{2}\alpha\mu_x \sin \tau \\ v_y = \frac{1}{2}\alpha\mu_y \cos \tau \\ \rho = 0. \end{aligned} \right\} \text{at } x=0 \left. \vphantom{\begin{aligned} V(x, b, t) = V(0, y, t) = 0 \\ v_x = \frac{1}{2}\alpha\mu_x \sin \tau \\ v_y = \frac{1}{2}\alpha\mu_y \cos \tau \\ \rho = 0. \end{aligned}} \right\} \text{GAP.}$$

For a system with grids, we have correspondingly:

$$\left. \begin{aligned} V(x, b, t) = V(0, y, t) = 0 \\ v_x = \frac{1}{2}\alpha\mu \sin \tau \\ v_y = 0 \\ i_x = 0. \end{aligned} \right\} \text{at } x=0 \left. \vphantom{\begin{aligned} V(x, b, t) = V(0, y, t) = 0 \\ v_x = \frac{1}{2}\alpha\mu \sin \tau \\ v_y = 0 \\ i_x = 0. \end{aligned}} \right\} \text{GRID.}$$

The formulas for v_x and v_y for the gap system are obtained by integrating the equations of motion of an electron (moving in a charge free drift space) across the region in which the gap field is appreciable, and retaining only linear terms in α . The beam coupling coefficients μ_x and μ_y which appear in these formulas are essentially the Fourier transforms of the longitudinal and transverse quasistatic electric fields generated in the charge free drift space by the voltage across the gap. To derive the explicit expressions for μ_x and μ_y which are listed in Section 2, the components of electric field are first expressed as Fourier integrals in which the integrands contain a product of an exponential function of the longitudinal coordinate and a hyperbolic function of the transverse coordinate; the specific nature of these functions is determined by the fact that, subject to the inequality $\beta_0\eta' \ll 1$, the hf electric field in the drift space due to an hf gap is a quasistatic field and a solution of Laplace's equation. With the aid of suitable boundary conditions on the electric field components and the Fourier integral theorem, the remainder of the derivation may be carried through without difficulty. The procedure which is outlined here is taken from unpublished notes of W. W. Hansen. See also reference 2.

Inasmuch as the expressions for μ_x and μ_y which are listed in Section 2 have been derived for a charge free drift space, they may not be applied validly beyond a first-order perturbation calculation with ha/η as expansion parameter.

The boundary condition on the hf charge

density for the gap system is based on the fact that it may be shown that the hf charge density ρ vanishes everywhere (to within quadratic terms in α) when space charge effects² are neglected. The range of application of this boundary condition is again limited to a first-order perturbation calculation in terms of ha/η .

Finally, the expression for v_x for a system with grids may be derived by integrating the equation of motion of an electron moving across the buncher grid space and retaining only linear terms in α .

It is at this point desirable to consider in some detail what takes place at points near the surfaces $y = \pm a$ (the surfaces of the unbunched beam). It is evident from a qualitative consideration of the effect of the space charge perturbation arising from the non-uniform distribution of charge that the beam alternately expands and contracts, such that points near the surfaces $y = \pm a$ are alternately inside and outside the beam with consequent violent fluctuations of charge density at these points. A literal representation of this state of affairs would not lend itself to a perturbation procedure whose validity requires that all hf quantities vary continuously with the perturbation parameter ha/η . For our purpose it is sufficient to replace the fluctuating boundary by an hf surface charge density (with consequent longitudinal surface current density) on the surfaces $y = \pm a$.

Let $y(x, y_0, t)$ denote the transverse coordinate of an electron which enters the drift space at $y = y_0$. The surface charge density then equals

$$\rho_s(x, a, t) = y(x, a, t) - a. \quad (6)$$

Similarly, for the surface current density we obtain

$$i_s(x, a, t) = y(x, a, t) - a. \quad (7)$$

There are analogous expressions for the surface charge density and surface current density for the boundary at $y = -a$. However, in view of the perfect symmetry which exists with respect to the surface $y = 0$, we shall in all future discussions confine our attention to the surface at $y = a$.

Having introduced a surface charge density, we require two additional boundary conditions. First, the potential is continuous at the boundary

of the beam, i.e.,

$$V_+(x, a, t) = V_-(x, a, t), \quad (8)$$

where the subscripts + and - designate the approach to $y=a$ from without and within the beam, respectively. Second, there is a discontinuity in the transverse component of the electric displacement with the amount of discontinuity determined by the surface charge density. In ordinary units, this relation is

$$(\partial V / \partial y)_- - (\partial V / \partial y)_+ = \rho_s / \epsilon_0.$$

In our reduced units, we get

$$(\partial V / \partial y)_- - (\partial V / \partial y)_+ = h^2 \rho_s. \quad (9)$$

5. CLASSIFICATION OF SOLUTIONS

Before actually applying the boundary conditions just outlined, we examine the consequences of the dynamical and field Eqs. (3)-(5). It is first convenient to remove the factor $\exp j(\omega t - k'x)$ from all hf quantities. If we then put for each hf quantity in (3)-(5) an expression of the form

$$f(x, y, t) = f(x, y) \exp j(\omega t - k'x),$$

the field equations and equations of motion may be restated to give

$$\{(\partial / \partial x - jk')^2 + \partial^2 / \partial y^2\} V = -h^2 \rho, \quad (3')$$

$$(\partial / \partial x - jk')v_z + \partial v_y / \partial y + \partial \rho / \partial x = 0, \quad (4')$$

$$\left. \begin{aligned} \partial v_z / \partial x &= -(\partial / \partial x - jk')V, \\ \partial v_y / \partial x &= -\partial V / \partial y. \end{aligned} \right\} \quad (5')$$

In the derivation of (5') we have used the approximation

$$\begin{aligned} d/dt &= \partial / \partial t + v_z \partial / \partial x + v_y \partial / \partial y \\ &\cong \partial / \partial t + v_0 \partial / \partial x. \end{aligned}$$

Now, differentiating (4') with respect to x and using (5'), we get

$$\{(\partial / \partial x - jk')^2 + \partial^2 / \partial y^2\} V = \partial^2 \rho / \partial x^2, \quad (10)$$

whence, comparing (3') and (10), we have

$$(\partial^2 / \partial x^2 + h^2) \rho = 0. \quad (11)$$

A consideration of (11) shows that there are two possible kinds of motion which may take place. In the first, there exists an hf charge density (i.e., the flow is non-solenoidal) with the

consequent implication that all hf quantities (V , ρ , v_z , v_y , i_z , i_y) have the general form

$$f_{12}(x, y) = \exp(-jh x) f_1(y) + \exp(jh x) f_2(y). \quad (12)$$

The non-solenoidal solution thus contributes two traveling waves, with the quantity h entering here as a debunching number.

In the second, the hf volume charge density vanishes identically (i.e., the flow is solenoidal). Then, both (3') and (10) degenerate into the same equation, viz., Laplace's equation. In fact, any potential function V which is a solution of Laplace's equation will give rise to a solenoidal type of motion. In what follows, we shall use the subscript 3 to denote solenoidal solutions.

It is important to note that, in Eq. (12) and the statement immediately preceding it, we have employed the term "non-solenoidal solution" in a special sense; evidently the motion which results from the addition of any arbitrary solenoidal solution to a non-solenoidal solution which is specified by (12) is non-solenoidal in the general sense of the word but is no longer characterized by (12) (except for the hf charge density ρ). However, for the purpose of classifying solutions, any general non-solenoidal motion may be considered as made up of a solenoidal flow plus a non-solenoidal flow which takes the form of Eq. (12). It is with this restricted meaning that we speak of a non-solenoidal solution throughout this paper.

In the following two sections, we discuss the general nature of the solenoidal and non-solenoidal solutions which may be used to describe an electron beam with elongated rectangular cross section, subject only to boundary conditions (8) and (9), with the additional requirement that $V(x, b) = 0$. Finally, in the last two sections we seek linear combinations of both types of solutions so as to satisfy the remaining boundary conditions whose nature is determined by the type of coupling between the beam and the modulating field.

6. THE SOLENOIDAL SOLUTION

For the solenoidal type of motion, the potential function V_3 must satisfy the following (Laplace's) equation:

$$\{(\partial / \partial x - jk')^2 + \partial^2 / \partial y^2\} V_3 = 0. \quad (13)$$

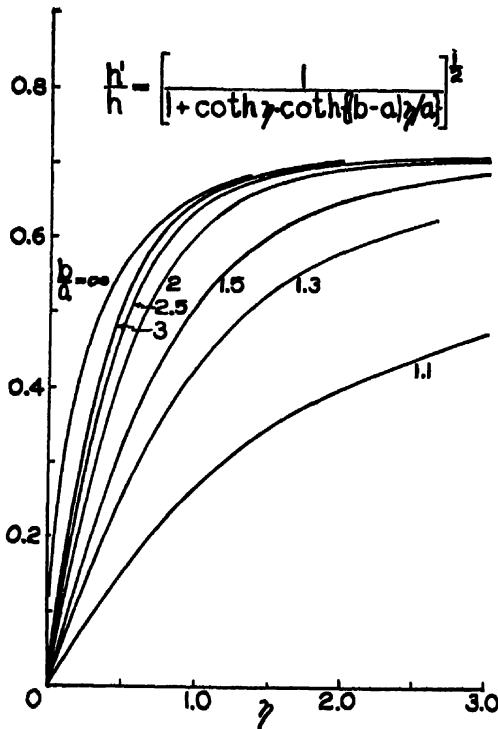


FIG. 2. The quantity h'/h as a function of η and b/a ; h' is the leading term in the expansions for both h_+ and h_- (the debunching numbers arising from the solenoidal solution).

Separating variables in (13), we may write for the potential function in the beam ($-a \leq y \leq a$)

$$V_s(x, y) = \sum_{\sigma} V_s^{(\sigma)} \cosh(u_{\sigma} y/a) / \cosh u_{\sigma} \cdot \exp j(\eta - u_{\sigma})x/a. \quad (14)$$

The dependence on the transverse coordinate y is chosen so as to make V_s an even function of y . Equations (5') suggest that we try similar type expressions for the hf velocities, viz.,

$$v_{zs}(x, y) = \sum_{\sigma} v_{zs}^{(\sigma)} \cdot \cosh(u_{\sigma} y/a) / \cosh u_{\sigma} \cdot \exp j(\eta - u_{\sigma})x/a, \quad (15)$$

$$v_{ys}(x, y) = \sum_{\sigma} v_{ys}^{(\sigma)} \cdot \sinh(u_{\sigma} y/a) / \cosh u_{\sigma} \cdot \exp j(\eta - u_{\sigma})x/a, \quad (16)$$

with the algebraic relations

$$v_{zs}^{(\sigma)} = v_{ys}^{(\sigma)} / j = u_{\sigma} / (\eta - u_{\sigma}) \cdot V_s^{(\sigma)}. \quad (17)$$

Outside the beam, for $a \leq y \leq b$, Eq. (8) plus the requirement that the potential function shall

vanish at the drift tube suggest that we write

$$V_s^*(x, y) = \sum_{\sigma} V_s^{(\sigma)} \cdot \sinh\{(b-y)u_{\sigma}/a\} / \sinh\{(b-a)u_{\sigma}/a\} \cdot \exp j(\eta - u_{\sigma})x/a. \quad (18)$$

It is clear that (18) is a solution of (13) and satisfies the boundary conditions

$$V_s^*(x, b) = 0; \quad V_s^*(x, a) = V_s(x, a). \quad (19)$$

For the determination of the u_{σ} 's we use boundary condition (9). Now the transverse displacement $y_s(x, a, t) - a$ may be expressed as the time integral of $v_{ys}(x, a, t)$ (in ordinary units) along the orbit of the electron; in reduced units this may be written

$$y_s(x, a, t) - a = v_0 \int_{\tau/a}^t v_{ys}\{x(t'), a, t'\} dt'. \quad (20)$$

If we carry out the integration along the undisturbed orbit of the electron in place of its actual orbit, we may then consider the factor $\exp j(\omega t - k'x)$ as a constant of the motion so that Eq. (20) may be restated to give

$$y_s(x, a) = \int_0^x v_{ys}(x', a) dx', \quad (21)$$

where we have set $dt' = dx'/v_0$. Note that the amplitude factor $y_s(x, a)$ refers not to the transverse coordinate but to the transverse displacement from the surface $y=a$, i.e.,

$$y_s(x, a, t) - a = y_s(x, a) \exp j(\omega t - k'x).$$

Hence, using (16) and (21), we get for the transverse displacement

$$y_s(x, a)/a = \sum_{\sigma} v_{ys}^{(\sigma)} \tanh u_{\sigma} \cdot [\{\exp j(\eta - u_{\sigma})x/a\} - 1] / j(\eta - u_{\sigma}). \quad (22)$$

There is an additive term in (22) which is independent of x ; the significance of this term will

TABLE I. Several limiting cases for the function $W(\eta) = [1 + \coth \eta \coth \{(b-a)\eta/a\}]^{-1}$.

Condition	$W(\eta)$
$\eta \rightarrow 0, b/a$ finite	$(b-a)/a \cdot \eta^2$
$\eta \rightarrow \infty, (b-a)/a \geq 1$	$\frac{1}{2} \cdot \{1 - \exp(-2\eta)\}$
$\eta \rightarrow \infty, 1 \geq (b-a)/a \geq 0$	$\frac{1}{2} \cdot [1 - \exp\{-2\eta(b-a)/a\}]$
η finite, $b \rightarrow a$	$(b-a)/a \cdot \eta \tanh \eta$
η finite, $b/a \rightarrow \infty$	$(1 + \coth \eta)^{-1}$

be discussed following the analysis of the non-solenoidal solution.

Boundary conditions (8) and (9) must hold separately for each component of potential, where a component of potential is specified by a particular functional dependence on x . Substituting (14), (18), and (22) into (9) and equating the coefficients of $\exp j(\eta - u_\sigma)x/a$ to zero, we have

$$\coth \{(b-a)u_\sigma/a\} = -\tanh u_\sigma \cdot \{1 + (ha)^2/(j\eta - ju_\sigma)^2\}. \quad (23)$$

We seek solutions of (23) where we neglect terms of order $(ha/\eta)^2$. First take ju_σ to be a real quantity; then (23) reduces to

$$\cosh (u_\sigma b/a) \cong 0,$$

or

$$ju_\sigma b/a \cong (\sigma + \frac{1}{2})\pi, \quad \sigma = 0, 1, \dots \quad (24)$$

and at the same time

$$\sinh \{(b-y)u_\sigma/a\} / \sinh \{(b-a)u_\sigma/a\} \cong \cosh (u_\sigma y/a) / \cosh u_\sigma. \quad (25)$$

It is evident from a consideration of the potential terms which (24) and (25) contribute when substituted into (14) and (18) that these terms are damped waves which, subject to the inequality $(ha/\eta)^2 \ll 1$, are identical with the waves which may be set up in a charge free drift tube.

We can also get solutions for u_σ out of (23) for u_σ real. Thus, we restate (23) in the form

$$u_\pm = \eta \pm h_\pm a, \quad (26)$$

with

$$h_\pm = h[1 + \coth u_\pm \cdot \coth \{(b-a)u_\pm/a\}]^{-1}, \quad (27)$$

where the $\pm$ replaces the σ subscript.

The potential terms contributed to (14) and (18) by (26) are obviously traveling waves, with the quantities h_+ and h_- serving as debunching numbers. There are hence a total of three such numbers, h , h_+ , and h_- , of which only the first appears in the theory of an unlimited plane beam.

We observe that under the usual restriction, $(ha/\eta)^2 \ll 1$,

$$h_\pm \cong h \cong h' \cong h[1 + \coth \eta \cdot \coth \{(b-a)\eta/a\}]^{-1} \quad (28)$$

and

$$|u_\pm - \eta| \ll \eta. \quad (29)$$

We digress briefly to consider the behavior of the function $W(\eta)$ where we define the latter as follows:

$$h' = h \{W(\eta)\}^{\frac{1}{2}} \quad (30)$$

or

$$W(\eta) = [1 + \coth \eta \cdot \coth \{(b-a)\eta/a\}]^{-1}. \quad (31)$$

Table I lists several limiting expressions for this function. Figure 2 contains a plot of the quantity h'/h as a function of η and b/a .

We list the expressions for V_z , v_{zs} , v_{ys} , and y_z obtained from Eqs. (14)–(16) and (22), respectively, upon substituting the values of u_σ given in (24) and (26); we use the notation $P_\sigma = (\sigma + \frac{1}{2})\pi$.

$$V_z(x, y) = V_{z+} \cdot \cosh (u_+ y/a) / \cosh u_+ \cdot \exp (-jh_+ x) + V_{z-} \cdot \cosh (u_- y/a) / \cosh u_- \cdot \exp (jh_- x) \\ + \exp (jk'x) \sum_{\sigma=0}^{\infty} V_{zs} \cdot \cos (P_\sigma y/b) / \cos (P_\sigma a/b) \cdot \exp (-P_\sigma x/b), \quad (32)$$

$$v_{zs}(x, y) = -u_+/h_+ a \cdot V_{z+} \cdot \cosh (u_+ y/a) / \cosh u_+ \cdot \exp (-jh_+ x) \\ + u_-/h_- a \cdot V_{z-} \cdot \cosh (u_- y/a) / \cosh u_- \cdot \exp (jh_- x) \\ + \exp (jk'x) \sum_{\sigma=0}^{\infty} -jP_\sigma/(\eta' + jP_\sigma) \cdot V_{zs} \cdot \cos (P_\sigma y/b) / \cos (P_\sigma a/b) \cdot \exp (-P_\sigma x/b), \quad (33)$$

$$v_{ys}(x, y) = -ju_+/h_+ a \cdot V_{z+} \cdot \sinh (u_+ y/a) / \cosh u_+ \cdot \exp (-jh_+ x) \\ + ju_-/h_- a \cdot V_{z-} \cdot \sinh (u_- y/a) / \cosh u_- \cdot \exp (jh_- x) \\ + \exp (jk'x) \sum_{\sigma=0}^{\infty} -jP_\sigma/(\eta' + jP_\sigma) \cdot V_{zs} \cdot \sin (P_\sigma y/b) / \cos (P_\sigma a/b) \cdot \exp (-P_\sigma x/b), \quad (34)$$

$$y_z(x, a)/a = u_+/(h_+ a)^2 \cdot V_{z+} \tanh u_+ \cdot \{\exp (-jh_+ x) - 1\} + u_-/(h_- a)^2 \cdot V_{z-} \tanh u_- \cdot \{\exp (jh_- x) - 1\} \\ + \sum_{\sigma=0}^{\infty} (-P_\sigma b/a)/(\eta' + jP_\sigma)^2 \cdot V_{zs} \tan (P_\sigma a/b) \cdot \{\exp (jk' - P_\sigma/b)x - 1\}. \quad (35)$$

The formula for $V_s^*(x, y)$ may be obtained directly from Eq. (32) by writing

$$\sinh \{(b-y)u_s/a\} / \sinh \{(b-a)u_s/a\}$$

for $\cosh (u_s y/a) / \cosh u_s$.

Summarizing this section, the application of boundary condition (9) has yielded two traveling solenoidal waves plus an infinite number of damped waves. The function of the former is to contribute to the propagation of the velocity modulation; the latter help satisfy the boundary conditions at $x=0$, particularly the condition $V(0, y, t)=0$.

7. THE NON-SOLENOIDAL SOLUTION

Our point of departure in the consideration of the non-solenoidal solution is once again boundary conditions (8) and (9). Making use of the functional dependence of all non-solenoidal hf quantities (as given in (12)) in conjunction with Eq. (5'), we get

$$\left. \begin{aligned} v_{z1} &= -(1+k'/h) V_1, \\ v_{z2} &= -(1-k'/h) V_2, \\ v_{y1} &= 1/jh \cdot \partial V_1 / \partial y, \\ v_{y2} &= -1/jh \cdot \partial V_2 / \partial y. \end{aligned} \right\} (36)$$

These relations together with an equation corresponding to (21) lead to the following expression for the transverse displacement function for the non-solenoidal case:

$$y_{12}(x, a) = 1/h^2 \cdot [(\partial V_1 / \partial y)_a \{\exp(-jh x) - 1\} + (\partial V_2 / \partial y)_a \{\exp(jh x) - 1\}]. \quad (37)$$

We defer the discussion of the additive term (independent of x) appearing in (37), which is the analog of the corresponding additive term in (22), until the close of this section.

Now in the space between the beam and the drift tube ($a \leq y \leq b$), the potential function $V_{12}^*(x, y)$ must have the form of Eq. (12) and simultaneously be a solution of Laplace's equation (13). We then get the following expression:

$$V_{12}^*(x, y) = V_1(a) \cdot \sinh \{(k'+h)(b-y)\} / \sinh \{(k'+h)(b-a)\} \cdot \exp(-jh x) + V_2(a) \cdot \sinh \{(k'-h)(b-y)\} / \sinh \{(k'-h)(b-a)\} \cdot \exp(jh x). \quad (38)$$

The form of the functional dependence of $V_{12}^*(x, y)$ on the transverse coordinate y is suggested by Eq. (18).

Applying boundary condition (9), we obtain, in view of Eq. (37),

$$(\partial V_1^* / \partial y)_a = (\partial V_2^* / \partial y)_a = 0, \quad (39)$$

where we have ignored, for the moment, the constant term in (37). The immediate consequence of (38) and (39) is that

$$\left. \begin{aligned} V_1(a) &= V_2(a) = 0, \\ \text{whence} \quad V_1^*(y) &= V_2^*(y) = 0, \end{aligned} \right\} (40)$$

or $V_{12}^*(x, y)$ vanishes identically.

In other words, the field outside the beam for the non-solenoidal type of motion is zero. The hf surface charge is such as to neutralize exactly the field due to the hf volume charge at all external points, i.e., the surface charge distribution is equal but opposite in sign to the Green's equivalent stratum corresponding to the volume charge distribution.⁵ Equation (40) is quite reasonable since the use of (12) and (13) requires that $V_{12}^*(x, y)$ be both non-solenoidal and solenoidal at the same time; this is clearly a contradiction which is resolved by the identical vanishing of $V_{12}^*(x, y)$.

We now rewrite Eq. (3') to read

$$\{\partial^2 / \partial y^2 - (k' + h)^2\} V_1 = -h^2 \rho_1 \quad (41)$$

plus a similar relation obtained from (41) by putting the subscript 2 in place of 1, and $-h$ for h .

It remains to guess at a suitable form for V_1 and V_2 , whence from (41) we will have expressions for ρ_1 and ρ_2 . Toward this end we are guided first, by the nature of the expressions obtained for the potential due to the traveling solenoidal waves (the first two terms on the right-hand side of Eq. (32)) and second, by our use of the inequality $(ha/\eta)^2 \ll 1$ as the basis for a perturbation procedure. We, therefore, try

$$V_1 = \{h/(k' + h)\}^2 \rho_1^0 + \sum_{\sigma=0}^{\infty} V_{1,\sigma+1} \cdot 1/\sigma! \cdot \partial^\sigma / \partial \eta^\sigma \cosh(\eta y/a) / \cosh \eta, \quad (42)$$

whereupon

$$\begin{aligned} \rho_1 &= \rho_1^0 - 1/(ha)^2 \\ &\cdot \sum_{\sigma=0}^{\infty} [V_{1,\sigma+1} + 2\eta V_{1,\sigma+1} - \{2\eta ha + (ha)^2\} V_{1,\sigma+1}] \\ &\cdot 1/\sigma! \cdot \partial^\sigma / \partial \eta^\sigma \cosh(\eta y/a) / \cosh \eta. \end{aligned} \quad (43)$$

⁵ Jeans, reference 3, p. 182.

Corresponding relations for V_2 and ρ_2 are got from (42) and (43) by substituting the subscript 2 for the subscript 1, and $-h$ for h .

Using (40), Eq. (42) implies that

$$\left. \begin{aligned} \rho_1^0 &= -(1 + \eta/ha)^2 V_{1,1}, \\ \rho_2^0 &= -(1 - \eta/ha)^2 V_{2,1}. \end{aligned} \right\} \quad (44)$$

We pause briefly to discuss the significance of the constant term appearing in the expression for y_{12} (given by (37)) as well as that of the analogous term in the corresponding equation for y_3 (given by (22)). As already pointed out in the discussion immediately preceding Eq. (23), boundary conditions (8) and (9) must hold separately for each component of potential, where a component of potential is specified by a particular functional dependence on x . It is then a direct consequence of Eqs. (8) and (9) that the constant term which appears in the total transverse displacement function $y_3 + y_{12}$ must vanish. This requirement (to which we shall refer as the consistency condition) follows since the constant term would otherwise contribute a traveling wave moving with beam velocity to the transverse displacement function, but there are no such terms in any of the other hf quantities (cf. Eqs. (12), (32) through (35)), and in particular in the expressions for the potentials or their transverse derivatives.

The problem which remains is to combine the solenoidal and non-solenoidal solutions so as to satisfy the consistency condition plus the boundary conditions (at $x=0$) for the gap and grid coupled systems, respectively. A complete solution for either case, therefore, requires the determination of explicit formulas for the potential coefficients V_{3+} , V_{3-} , V_{2+} , $V_{1, \sigma+1}$, $V_{2, \sigma-1}$ ($\sigma=0, 1, \dots$) to within a sufficient degree of accuracy so as to enable the calculation of the leading term and the term of order ha/η relative to the leading term for each of the hf quantities (v_x , v_y , ρ , $y(x, a)$) which are in turn needed to evaluate the leading and first-order terms for the longitudinal volume and surface current densities.

We now assume for both the gap and grid systems that the coefficients $V_{3\sigma}$, arising from the infinite set of damped solenoidal waves, are all of the order $(ha/\eta)^2$ or higher. With this hypothesis, it is found possible to satisfy the consistency condition as well as the boundary conditions on the entering hf components of the longitudinal and transverse velocities and the entering hf component of beam current density (or charge density), and so evaluate V_{3+} , V_{3-} , $V_{1, \sigma+1}$, $V_{2, \sigma+1}$ explicitly. The remaining requirement, *viz.*, that the potential vanish at $x=0$, leads to two conditions on the potential coefficients, the one holding outside the beam, and the other within the beam. These conditions not only verify the assumption on the $V_{3\sigma}$'s but also provide a means for their determination correct to terms in $(ha/\eta)^2$, although we have no actual need of them.

8. GAP COUPLING

The absence of an hf charge density at $x=0$ makes it possible to satisfy all the boundary conditions at the entering plane as well as the consistency condition without using the non-solenoidal waves. In discarding the latter, we are implicitly assuming that the contributions to the hf components of the longitudinal and transverse velocities by the non-solenoidal waves are of the order $(ha/\eta)^2$ relative to the leading terms in the hf velocities. We verify these statements by showing that it is indeed possible to construct a solution which will satisfy the boundary conditions and the consistency condition by simply using the two traveling solenoidal waves. In view of the remarks at the end of Section 7, we also neglect the infinite sum of damped waves.

Equations (9) and (35) then yield for the consistency condition

$$u_+/(h_+a)^2 \cdot V_{3+} \tanh u_+ + u_-/(h_-a)^2 \cdot V_{3-} \tanh u_- \cong 0. \quad (45)$$

The boundary conditions on the entering hf velocity components assume the form (making use of (33) and (34))

$$\alpha\mu/2j \cdot \cosh(\eta y/a)/\cosh \eta' \cong -u_+/h_+a \cdot V_{3+} \cdot \cosh(u_+y/a)/\cosh u_+ + u_-/h_-a \cdot V_{3-} \cdot \cosh(u_-y/a)/\cosh u_-, \quad (46)$$

$$\alpha\mu/2j \cdot \sinh(\eta y/a)/\cosh \eta' \cong -u_+/h_+a \cdot V_{3+} \cdot \sinh(u_+y/a)/\cosh u_+ + u_-/h_-a \cdot V_{3-} \cdot \sinh(u_-y/a)/\cosh u_-. \quad (47)$$

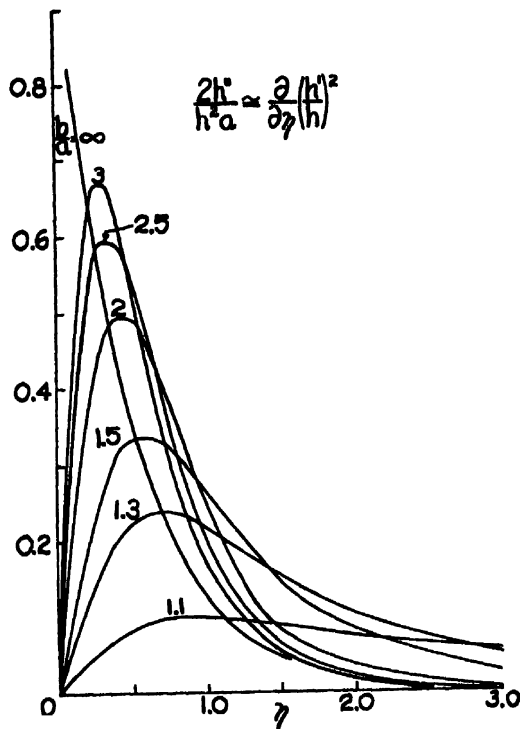


FIG. 3. The quantity $2h''/h'a$ as a function of η and b/a ; h'' and $-h''$ are the first correction terms in the expansions for h_+ and h_- , respectively (where h_+ and h_- are the debunching numbers arising from the solenoidal solution).

In (45)–(47) we have neglected terms of order $(ha/\eta)^2$ relative to the leading terms.

If we expand the quantities $\cosh(u_{\pm}y/a)$ and $\sinh(u_{\pm}y/a)$ into Taylor series about the point η , and make use of the substitutions

$$f(u_+) + f(u_-) = 2f(\eta) \{1 + O((ha/\eta)^2)\}$$

and

$$f(u_+) - f(u_-) = 2h'a \cdot \partial/\partial\eta f(\eta) \cdot \{1 + O(ha/\eta)\},$$

$$v_z/v_0 \cong \alpha\mu/2j \cdot \sinh \eta / \cosh \eta' \cdot \exp(-jh''x) \cdot [\cos(h'x) \cdot \cosh(\eta y/a) / \sinh \eta - j \sin(h'x) \cdot \partial/\partial\eta \{h'a \cdot \cosh(\eta y/a) / \sinh \eta\}], \quad (51)$$

$$v_y/v_0 \cong \alpha\mu/2 \cdot \sinh \eta / \cosh \eta' \cdot \exp(-jh''x) \cdot [\cos(h'x) \cdot \sinh(\eta y/a) / \sinh \eta - j \sin(h'x) \cdot \partial/\partial\eta \{h'a \cdot \sinh(\eta y/a) / \sinh \eta\}], \quad (52)$$

$$y(x, a) \cong \alpha\mu/2 \cdot \sinh \eta / \cosh \eta' \cdot \exp(-jh''x) \cdot \sin(h'x)/h'. \quad (53)$$

In (51)–(53) terms of order $(ha/\eta)^2$ with respect to the leading terms are neglected. There is no zero-order term in (53) because of the consistency condition.

To calculate the volume and surface current densities, we need simply put $i_s = \rho_0 v_z$ and

then (46) and (47) are not independent relations but reduce to the same equation, *viz.*,

$$\alpha\mu/2j \cdot 1/\cosh \eta' \cong -u_+/h_+a \cdot V_{z+} \cdot 1/\cosh u_+ + u_-/h_-a \cdot V_{z-} \cdot 1/\cosh u_-, \quad (48)$$

where we neglect terms of order $(ha/\eta)^2$.

Equations (45) and (48) possess the solution

$$-u_{\pm}/h_{\pm}a \cdot V_{z\pm} = \pm \alpha\mu/4j \cdot h_{\pm}/h' \cdot \sinh \eta / \cosh \eta' \cdot \coth u_{\pm} \cdot \{1 + O_{\pm}((ha/\eta)^2)\}, \quad (49)$$

where O_+ and O_- are remainder functions with leading terms of the second order in ha/η .

It follows from (36) that the hf components of potential contributed by the two non-solenoidal waves are of the third order in ha/η , whence the total potential function at $x=0$ will equal a known function (quadratic in ha/η) arising from the traveling solenoidal waves plus an infinite sum of local damped waves with coefficients V_{zs} . Inasmuch as $V(0, y) = V^*(0, y) = 0$, all of the coefficients V_{zs} are quadratic in ha/η . We have, therefore, verified our initial assumption about the order of magnitude of the V_{zs} 's.

It is now a straightforward procedure to compute the total hf longitudinal and transverse velocities, the transverse displacement function, and finally the effective beam volume and surface currents. For the remainder of this section, we state all quantities in ordinary units.

It will prove convenient to use the notation

$$2h'' = h_+ - h_- \cong h^2 a \cdot \partial/\partial\eta (h'/h)^2. \quad (50)$$

The quantity $2h''/h'a$ is plotted in Fig. 3 as a function of η and b/a .

Inserting (49) into (33), (34) and (35), we get

have

$$I_s/I_0 = \langle v_z/v_0 \cdot \mu \cosh(\eta y/a) / \cosh \eta' \rangle \times \exp j(\omega t - k'x) \\ \cong \alpha \mu^2 / 2j \cdot \left[\frac{1}{2} \{ 1 + \sinh(2\eta)/2\eta \} \cdot 1/\cosh^2 \eta' \right] \cdot \cos(h'x) \cdot \exp j\{\omega t - (k' + h'')x\}. \quad (54)$$

In similar fashion, we write for the effective surface current

$$I_s/I_0 = \mu \cosh \eta / \cosh \eta' \cdot y(x, a) / a \cdot \exp j(\omega t - k'x) \\ \cong \alpha \mu^2 / 2 \cdot \{ \sinh(2\eta)/2\eta \cdot 1/\cosh^2 \eta' \} \cdot k'/h' \cdot \sin(h'x) \cdot \exp j\{\omega t - (k' + h'')x\}. \quad (55)$$

The quantity in the curly brackets in Eq. (55) is plotted in Fig. 4 as a function of η and b/a .

In deriving Eqs. (54) and (55), we have retained leading terms only. Hence the first neglected term in I_s is of the order ha/η ; the first neglected term in I_s is also of the order ha/η since the transverse displacement function (53) has no zero-order term.

Comparing the relative orders of magnitude of I_s and I_v , we note that the hyperbolic function enclosed in the brackets in (55) is always larger than the corresponding function in (54) for $\eta > 0$ and equal to it for $\eta = 0$; this statement follows readily from the inequality $\sinh(2\eta)/2\eta \geq 1$ for $\eta \geq 0$. Furthermore, I_s is of the order h'/k' (a quantity smaller than ha/η) with respect to I_s , and our entire procedure has been based on the assumption that $(ha/\eta)^2 \ll 1$.

The total hf current is the sum of the surface and volume currents. Noting that the leading terms in I_s and I_v are 90° out of phase, it follows from the preceding two paragraphs that the amplitude factor for the total current, which equals $(|I_s|^2 + |I_v|^2)^{1/2}$, contains a leading term of order $(k'/h')^2$ contributed by the surface current, with the following term of zero order, contributed by both the volume and surface currents. The volume current is therefore insignificant except in the immediate vicinity of zeros of $\sin(h'x)$.

Retaining only the surface current, maximum hf current occurs for a drift tube length determined by the first maximum of $\sin(h'x)$; this length corresponds to maximum gain for a single stage amplifier at low input levels. Hence

$$x_{\max} = \pi/2h'. \quad (56)$$

The perturbation procedure is no longer valid when the transverse displacement (53) exceeds $b-a$. When this happens, current flows directly to the drift tube in hf pulses while at the same time the hf beam current drops considerably. This effect should become sharp for $(b-a)/a \ll 1$, i.e., there will exist a critical input level such that below this level the system will behave as an amplifier with a given "small signal" voltage or power gain, but above this level the output will remain fairly constant.

The critical input level is determined by the relation

$$|y(x, a)| = b - a. \quad (57)$$

For a tube designed to give a maximum gain at low input levels, (53) and (57) yield

$$\alpha = 2ha'\mu \cdot (\eta \coth \eta)^{1/2} \cdot \{(b-a)/a\}^{1/2}, \quad (58)$$

where we have assumed $(b-a)/a \ll 1$ together with the expression for h' given in Table I for this condition.

9. GRID COUPLING

We return once more to the use of reduced units. Our input conditions on the longitudinal current density and velocity are $i_z = \rho + v_z = 0$ and $v_z = \alpha\mu/2j \cdot \exp(j\omega t)$; so that, in view of (12),

$$\rho_1 + \rho_2 = -\alpha\mu/2j. \quad (59)$$

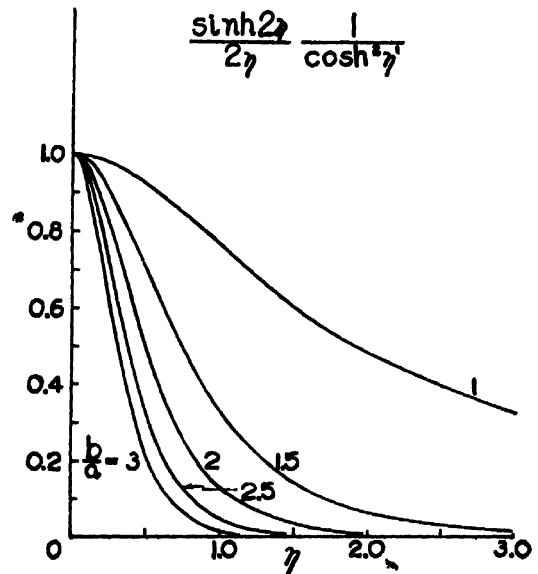


FIG. 4. A coefficient required to evaluate the effective surface current with gap coupling (refer to Eq. (55)).

In terms of (43) and (44), Eq. (59) yields the following two relations:

$$\{1 + (\eta/ha)^2\} \cdot (V_{1,1} + V_{2,1}) + 2\eta/ha \cdot (V_{1,1} - V_{2,1}) = \alpha\mu/2j, \quad (60)$$

$$2\eta/ha \cdot (V_{1,\sigma+1} - V_{2,\sigma+1}) = -(V_{1,\sigma+1} + V_{2,\sigma+1}) + 2\eta/(ha)^2 \cdot (V_{1,\sigma+2} + V_{2,\sigma+2}) + 1/(ha)^2 \cdot (V_{1,\sigma+3} + V_{2,\sigma+3}), \quad \sigma = 0, 1, \dots \quad (61)$$

We again make the hypothesis that the potential coefficients $V_{3\sigma}$, which specify the damped local solenoidal waves, are all of the order $(ha/\eta)^2$. This means that in the statement of the boundary conditions on the entering longitudinal and transverse velocities and of the consistency condition, terms due to these local waves may be omitted. This hypothesis will be verified along the lines given in the discussion toward the end of Section 7.

With the help of Eqs. (12), (33), (36), (42), and (44), the boundary condition on the longitudinal velocity at $x=0$ yields the following relation:

$$\begin{aligned} & \{(V_{1,1} + V_{2,1}) + \eta/ha \cdot (V_{1,1} - V_{2,1})\} \\ & - \sum_{\sigma=0}^{\infty} \{(V_{1,\sigma+1} + V_{2,\sigma+1}) + \eta/ha \cdot (V_{1,\sigma+1} - V_{2,\sigma+1})\} \\ & \cdot 1/\sigma! \cdot \partial^\sigma / \partial \eta^\sigma \cosh(\eta y/a) / \cosh \eta \\ & - u_+/h_+a \cdot V_{3+} \cdot \cosh(u_+y/a) / \cosh u_+ \\ & + u_-/h_-a \cdot V_{3-} \cdot \cosh(u_-y/a) / \cosh u_- \\ & \cong \alpha\mu/2j. \quad (62) \end{aligned}$$

Setting up the corresponding boundary condition on the transverse velocity at $x=0$, where we now employ Eq. (34) in place of (33) in the above listing, we get, following a single differentiation with respect to y plus a rearrangement of terms,

$$\begin{aligned} & \sum_{\sigma=0}^{\infty} \{\eta/ha \cdot (V_{1,\sigma+1} - V_{2,\sigma+1}) + 2/ha \cdot (V_{1,\sigma+2} - V_{2,\sigma+2}) \\ & + 1/\eta ha \cdot (V_{1,\sigma+3} - V_{2,\sigma+3})\} \cdot 1/\sigma! \\ & \cdot \partial^\sigma / \partial \eta^\sigma \cosh(\eta y/a) / \cosh \eta + u_+/h_+a \cdot V_{3+} \cdot \cosh(u_+y/a) / \cosh u_+ \\ & - u_-/h_-a \cdot V_{3-} \cdot \cosh(u_-y/a) / \cosh u_- \cong 0. \quad (63) \end{aligned}$$

Finally, using (35) and (37) and applying the consistency condition, we obtain the relation

$$\begin{aligned} & 1/(ha)^2 \cdot \sum_{\sigma=0}^{\infty} \{\eta(V_{1,\sigma+1} + V_{2,\sigma+1}) \\ & + (V_{1,\sigma+2} + V_{2,\sigma+2})\} \cdot 1/\sigma! \cdot \partial^\sigma / \partial \eta^\sigma \tanh \eta \\ & + u_+/(h_+a)^2 \cdot V_{3+} \cdot \tanh u_+ \\ & + u_-/(h_-a)^2 \cdot V_{3-} \cdot \tanh u_- \cong 0. \quad (64) \end{aligned}$$

Note that in Eqs. (62)–(64) additive terms of order $(ha/\eta)^2$ have been discarded.

The foregoing set of Eqs. ((60) through (64)) is adequate to evaluate explicitly the potential coefficients V_{3+} , V_{3-} , $V_{1,\sigma+1}$, $V_{2,\sigma+1}$ within an accuracy determined by our hypothesis on the order of magnitude of the coefficients $V_{3\sigma}$. The calculations required for this evaluation are given in reference 2. We list the results:

$$\begin{aligned} V_{1,1} + V_{2,1} &= -\alpha\mu/2j \cdot (ha/\eta)^2 / (1 - (ha/\eta)^2), \\ V_{1,1} - V_{2,1} &= \alpha\mu/2j \cdot (ha/\eta) / (1 - (ha/\eta)^2), \\ V_{1,2} + V_{2,2} &= \alpha\mu/2j \cdot (ha/\eta)^2 \eta \{1 + 0((ha/\eta)^2)\}, \\ V_{1,\sigma+3} + V_{2,\sigma+3} &= \alpha\mu/2j \cdot 0((ha/\eta)^4), \\ & \sigma = 0, 1, \dots, \quad (65) \\ V_{1,\sigma+2} - V_{2,\sigma+2} &= \alpha\mu/2j \cdot 0((ha/\eta)^2), \\ & \sigma = 0, 1, \dots, \end{aligned}$$

$$\begin{aligned} u_{\pm}/h_{\pm}a \cdot V_{3\pm} &= \mp \alpha\mu/4j \cdot \{1 \pm ha/\eta \cdot K \\ & + 0_{\pm}((ha/\eta)^2)\}, \\ ha/\eta \cdot K &= h''/h' \cdot (1 + 0(ha/\eta)). \end{aligned}$$

There is one final requirement which the potential coefficients must satisfy arising from the boundary condition on the potential at $x=0$, viz., $V(0, y) = V'(0, y) = 0$. Setting up the potential function within and without the beam with help of (12), (32), (42), and (44) together with the statement immediately following (35), and using the potential coefficients for the traveling solenoidal and non-solenoidal waves given in (65), we obtain a known function (quadratic in ha/η) plus an infinite sum of damped waves with coefficients $V_{3\sigma}$. Since the potential vanishes at $x=0$, all the coefficients $V_{3\sigma}$ are quadratic in ha/η , thus justifying our initial assumption.

Equation (65) constitutes a complete solution of the problem. We proceed to list some im-

portant hf quantities, returning once more to the use of ordinary units.

$$\rho/\rho_0 \cong \alpha\mu/2 \cdot \{k' \sin(hx)/h + j \cos(hx)\}, \quad (66)$$

$$v_z/v_0 \cong \alpha\mu/2j \cdot [\{\cosh(\eta y/a)/\cosh \eta\} \{\exp(-jh''x) \cdot \cos(h'x) - \cos(hx)\} + \cos(hx)], \quad (67)$$

$$y(x, a)/a \cong \alpha\mu/2 \cdot [(\tanh \eta/\eta) \{k' \exp(-jh''x) \cdot \sin(h'x)/h' - k' \sin(hx)/h\} + j(\partial/\partial \eta \tanh \eta) \{\exp(-jh''x) \cdot \cos(h'x) - \cos(hx)\}]. \quad (68)$$

The terms which are neglected in (66)–(68) are of the order ha/η .

To calculate the volume and surface current densities, we use the relations

$$i_x/i_0 = \rho/\rho_0 + v_z/v_0 \quad \text{and} \quad i_s/i_0 = y(x, a).$$

The total volume current then equals

$$I_v/I_0 = \langle i_x/i_0 \rangle$$

$$I_v/I_0 = I_v/I_0 + I_s/I_0$$

$$\cong \alpha\mu/2 \cdot [(\tanh \eta/\eta) \cdot k' \exp(-jh''x) \cdot \sin(h'x)/h' + (1 - \tanh \eta/\eta) \cdot k' \sin(hx)/h + j(\eta \partial/\partial \eta \tanh \eta/\eta) \{\exp(-jh''x) \cdot \cos(h'x) - \cos(hx)\}]. \quad (70)$$

The hyperbolic function coefficients appearing in (68)–(70) are plotted in Fig. 5.

The importance of the surface current in the

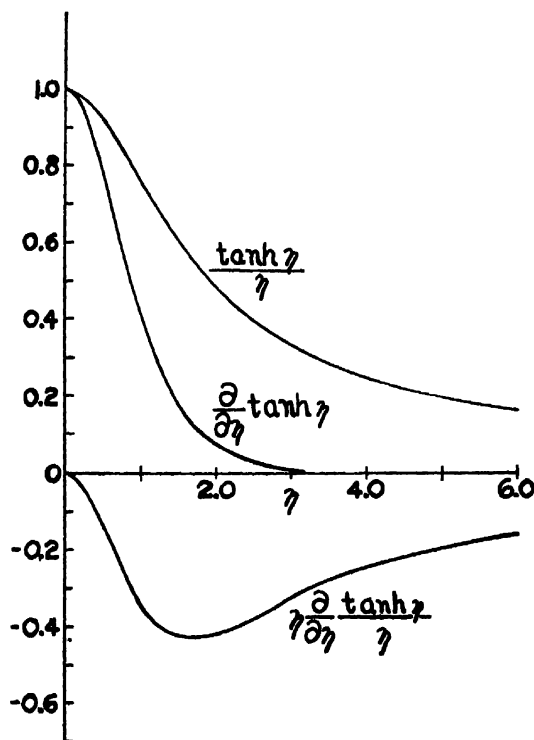


FIG. 5. Coefficients required to evaluate the volume, surface and total currents with grid coupling (refer to Eqs. (69)–(70)).

$$\cong \alpha\mu/2 \cdot [k' \sin(hx)/h - j(\tanh \eta/\eta) \cdot \{\exp(-jh''x) \cdot \cos(h'x) - \cos(hx)\}]. \quad (69)$$

For the total surface current, we obtain

$$I_s/I_0 = y(x, a)/a, \quad (68')$$

and it is hence given directly by Eq. (68).

We compute finally the total current:

usual design of klystron amplifier with grid coupling is established by Figs. 2 and 5. It is easily possible for the first term in Eq. (70) to exceed the second by a factor of three or four. Indeed, for $\eta < 1$, the first term in (70) is the only one of importance, and the motion becomes solenoidal.⁶ The length of drift tube for maximum hf current is then given by the relation which we obtained previously for the gap system, *vis.*,

$$x_{\max} = \pi/2h'. \quad (56)$$

The corresponding maximum of current has the value

$$(I_v/I_0)_{\max} \cong \alpha\mu k'/2h' = \alpha\mu/2 \cdot \eta/ha \cdot h/h'. \quad (71)$$

We note finally that, as $\eta \rightarrow \infty$, Eq. (70) reduces to the correct limiting expression for an unbounded beam,

$$(I_v/I_0)_{\eta \rightarrow \infty} \cong \alpha\mu/2 \cdot k' \sin(hx)/h. \quad (72)$$

⁶ R. H. Varian has shown, by a physical argument, that the flow must become solenoidal for a beam of circular cross section in the limiting case of extremely small radius (for a fixed total current). For a given amount of velocity modulation, it follows from Gauss's theorem in electrostatics that the longitudinal field due to space charge approaches a finite limiting value as the cross section is made smaller. At the same time, the radial field at the edge of the beam increases without limit (roughly inversely as the radius of the beam). Hence the bunching process is such as to tend toward expansion and contraction at constant density as the cross section is reduced. It is of interest to note that, in the case treated in this paper, we also approach a solenoidal flow⁶ as the transverse dimension of the beam is reduced for a fixed current density, although Varian's argument is no longer applicable

On Vibrations of Shallow Spherical Shells

ERIC REISSNER

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received July 30, 1946)

The problem of the axi-symmetrical vibrations of a shallow spherical shell is reduced to two simultaneous differential equations for the tangential and normal components of displacement. The solution of these equations is obtained in terms of Bessel functions. With this solution the third-order determinant for the frequencies of a shell segment with clamped edge is given but not evaluated. Instead, an approximate value for the lowest frequency is calculated by means of the Rayleigh-Ritz procedure [Eq. (40)]. It is found that very little curvature is needed to modify appreciably the corresponding flat-plate frequency. The approximations which are made are those customary for shallow shells: (1) omission of the transverse shear term in the tangential force equilibrium equation, and (2) relations between couples and changes of curvatures as in the theory of flat plates.

1. INTRODUCTION

IN this note we consider the problem of the rotationally symmetric vibrations of a thin, shallow, elastic, spherical shell. Our aim is to arrive at a formula which contains the modifying effect of the curvature of the shell on the corresponding formulas for the flat plate. We take the basic equations as in an earlier paper on the statics of shallow spherical shells.¹ These equations are reduced to two simultaneous equations for normal and tangential components of displacement. It is found that these two equations have a symmetrical structure of such a kind that their complete solution in terms of Bessel functions is possible. Determination of the frequencies of a shell segment with one boundary requires the evaluation of a third-order determinant having as its elements products of Bessel functions and rational functions. For a shell segment with two boundaries a corresponding sixth-order determinant occurs.

An upper bound for the lowest frequency of a shell segment with clamped edge is obtained by means of the Rayleigh-Ritz procedure. This formula shows the expected pronounced effect of even small curvature on the values of the frequency.

After the work of this paper had been completed it was found that similar results had been obtained earlier by K. Federhofer.² As our

method of derivation is quite different from Federhofer's, and as moreover Federhofer's work is published in a relatively inaccessible place, it is thought that publication of the present developments still may be of some interest.

2. FORMULATION OF THE PROBLEM

We take the equation of the middle surface of the shell in the form

$$z = (R^2 - r^2)^{1/2} - (R - h), \quad (1)$$

where R is the radius of the shell, h the distance between base plane and apex of the shell, and r the radius vector in the base plane (Fig. 1). Let $r = a$ be the intersection of the shell surface with the plane $z = 0$, so that according to (1),

$$a^2 = 2hR - h^2. \quad (2)$$

For a shallow shell we have the order of mag-

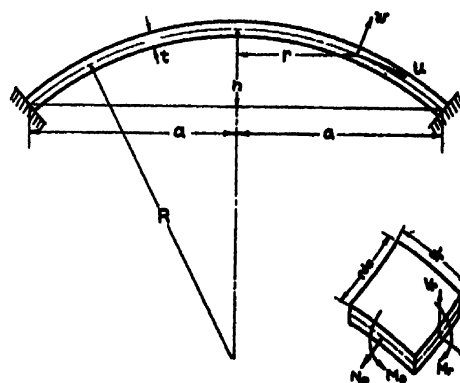


FIG. 1. Section of shallow spherical shell segment and element of shell, with indication of symbols used in text.

¹ E. Reissner, *J. Math. and Phys.* **25**, 80-85 (1946).

² K. Federhofer, *Sitz. Akad. d. Wiss. Wien* **146**, 57-69 (1937).

nitude relation

$$\frac{dz}{dr} = -\frac{r}{(R^2 - r^2)^{1/2}} \sim -\frac{r}{R} = o(1), \quad (3)$$

which must hold for all $r \leq a$. As a consequence of (3) we have that $h \ll R$, and Eq. (2) may be replaced by

$$a = (2hR)^{1/2}. \quad (4)$$

We restrict attention to rotationally symmetric vibrations with circular frequency ω and omit the time factor in the equations of the problem. We then have as equations of equilibrium

$$(rN_r)' - N_\theta + t\rho\omega^2 ru = 0, \quad (5a)$$

$$(rV_r)' - (r/R)(N_r + N_\theta) + t\rho\omega^2 rw = 0, \quad (5b)$$

and

$$(rM_r)' - M_\theta - rV_r = 0. \quad (5c)$$

In (5) primes denote differentiation with respect to r ; u and w are the meridional and normal component of displacement; N_r and N_θ are direct stress resultants; V_r is the transverse shear stress resultant; and M_r and M_θ are the bending stress

couples. Equation (5a) has been simplified by the omission of the transverse shear term.¹

We take as stress strain relations¹

$$N_r = C \left[u' + \frac{w}{R} + \nu \left(\frac{u}{r} + \frac{w}{R} \right) \right], \quad (6a)$$

$$N_\theta = C \left[\frac{u}{r} + \frac{w}{R} + \nu \left(u' + \frac{w}{R} \right) \right], \quad (6b)$$

$$M_r = -D \left(w'' + \frac{\nu}{r} w' \right), \quad (6c)$$

and

$$M_\theta = -D \left(\frac{w'}{r} + \nu w' \right), \quad (6d)$$

where

$$C = \frac{Et}{1 - \nu^2} \quad \text{and} \quad D = \frac{Et^3}{12(1 - \nu^2)}. \quad (7a, b)$$

From (6c, d) and (5c) we obtain for a shell of constant thickness t

$$V_r = -D(w'' + w'/r)', \quad (8)$$

thereby expressing all resultants and couples in terms of the components of displacement.

3. THE DIFFERENTIAL EQUATIONS FOR THE COMPONENTS OF DISPLACEMENT

We introduce (6a) and (6b) into (5a) and obtain

$$ru'' + u' - u/r + (1 + \nu)rw'/R + (t\rho\omega^2/C)ru = 0. \quad (9)$$

We next introduce (6a), (6b), and (8) into (5b) and obtain

$$[rw'' + w' - w'/r]' + ((1 + \nu)C/RD)[ru' + u + 2rw/R] - (t\rho\omega^2/D)rw = 0. \quad (10)$$

The system (9) and (10) which corresponds to one differential equation of the sixth order can be transformed in such a way that its complete solution is expressed in terms of Bessel functions.

To see this we use the identity

$$rf'' + f' - f/r \equiv r[r^{-1}(rf)']' \quad (11)$$

and introduce an operator L defined by

$$L(g) \equiv r[r^{-1}g']'. \quad (12)$$

With (11) and (12) we write (9) and (10) in the form

$$L(ru) + (1 + \nu)rw'/R + (t\rho\omega^2/C)ru = 0 \quad (13)$$

and

$$[L(rw')] + ((1 + \nu)C/RD)[(ru)' + 2rw/R] - (t\rho\omega^2/D)rw = 0. \quad (14)$$

In (13) and (14) the quantities u and w occur principally in the form ru and rw' . This suggests that we introduce two new dependent variables y and z defined as

$$y = ru \quad \text{and} \quad z = rw'. \quad (15a, b)$$

Furthermore, we divide (14) by r , differentiate the result, and then multiply by r . If this is done equations (13) and (14) become

$$L(y) + (t\rho\omega^2/C)y + (1+\nu)z/R = 0 \quad (16)$$

and

$$LL(z) - (t\rho\omega^2/D)z + ((1+\nu)C/RD)[L(y) + 2z/R] = 0. \quad (17)$$

When $R = \infty$, Eqs. (16) and (17) are equivalent to known equations for the uncoupled extensional and flexural vibrations, with no radial nodes, of circular plates.

We put as abbreviations in (16) and (17)

$$t\rho\omega^2/C = (1-\nu^2)\rho\omega^2/E = \alpha^2, \quad (18a)$$

$$t\rho\omega^2/D = 12(1-\nu^2)\rho\omega^2/Et^2 = k^4, \quad (18b)$$

$$(1+\nu)R = 2(1+\nu)h/a^2 = m_1^2, \quad (18c)$$

$$(1+\nu)C/RD = 24(1+\nu)h/a^2t^2 = m_2^2, \quad (18d)$$

$$2(1+\nu)C/R^2D = 96(1+\nu)h^2/a^4t^2 = k^4 - \beta^4 \quad (18e)$$

and have

$$L(y) + \alpha^2y + m_1^2z = 0 \quad (19)$$

and

$$LL(z) - \beta^4z + m_2^2L(y) = 0. \quad (20)$$

4. SOLUTION OF THE DIFFERENTIAL EQS. (19) AND (20)

We note that the equation

$$L(f) + \mu^2f = 0 \quad (21)$$

has the Bessel-function solution

$$f = rZ_1(\mu r). \quad (22)$$

We set in (19) and (20)

$$y = ArZ_1(\mu r), \quad \text{and} \quad z = BrZ_1(\mu r) \quad (23a, b)$$

and obtain

$$\begin{cases} (\alpha^2 - \mu^2)A + m_1^2B = 0 \\ -m_2^2\mu^2A + (\mu^4 - \beta^4)B = 0 \end{cases} \quad (24a, b)$$

The condition of vanishing determinant of (24) is

$$(\alpha^2 - \mu^2)(\mu^4 - \beta^4) + m_1^2m_2^2\mu^2 = 0. \quad (25a)$$

If we multiply (25a) by a^4 , where a may be the radius of the base of the shell segment, we obtain, with (18),

$$\left[(1-\nu^2)\frac{\rho\omega^2a^2}{E} - (\mu a)^2 \right] \left[(\mu a)^4 - 12(1-\nu^2)\frac{\rho\omega^2a^4}{Et^2} + 96(1+\nu)\frac{h^2}{t^2} \right] + 48(1+\nu)\frac{h^2}{t^2}(\mu a)^2 = 0. \quad (25b)$$

It follows from (25b) that the three roots μ_1 , μ_2 , and μ_3 of (25) are of the form

$$\mu a = f \left[\omega \frac{a^2}{t} \left(\frac{(1-\nu^2)}{E} \rho \right)^{\frac{1}{2}}, \frac{h}{t}, \frac{h}{a}, \nu \right]. \quad (26)$$

From (24a) follows for the amplitude ratios

$$A/B = m_1^2/(\alpha^2 - \mu^2). \quad (27)$$

Combining the solutions corresponding to different values of μ and taking J_1 as well as Y_1 as part

of Z_1 , we may write

$$z = rw' = r \{ B_1 J_1(\mu_1 r) + C_1 Y_1(\mu_1 r) + B_2 J_1(\mu_2 r) + C_2 Y_1(\mu_2 r) + B_3 J_1(\mu_3 r) + C_3 Y_1(\mu_3 r) \}, \quad (28)$$

and

$$y = ru = rm_1^2 \left\{ \frac{B_1 J_1(\mu_1 r)}{\alpha^2 - \mu_1^2} + \frac{C_1 Y_1(\mu_1 r)}{\alpha^2 - \mu_1^2} + \frac{B_2 J_1(\mu_2 r)}{\alpha^2 - \mu_2^2} + \frac{C_2 Y_1(\mu_2 r)}{\alpha^2 - \mu_2^2} + \frac{B_3 J_1(\mu_3 r)}{\alpha^2 - \mu_3^2} + \frac{C_3 Y_1(\mu_3 r)}{\alpha^2 - \mu_3^2} \right\}. \quad (29)$$

For an annular shell all six constants of integration are needed in order to satisfy three boundary conditions on each edge. For a complete shell segment with no singularity at the apex we must have $C_1 = C_2 = C_3 = 0$.

Integrating (27) we obtain for the deflection w ,

$$w = -B_1 J_0(\mu_1 r)/\mu_1 - C_1 Y_0(\mu_1 r)/\mu_1 - B_2 J_0(\mu_2 r)/\mu_2 - C_2 Y_0(\mu_2 r)/\mu_2 - B_3 J_0(\mu_3 r)/\mu_3 - C_3 Y_0(\mu_3 r)/\mu_3. \quad (30)$$

5. FREQUENCY EQUATION FOR SHELL SEGMENT WITH CLAMPED EDGE

As an example of the form which the equation determining the frequency assumes, we consider the motion of a shell determined by the conditions that there is no singularity for $r=0$ and that along the edge $r=a$ there is complete restraint,

$$w(a) = w'(a) = u(a) = 0. \quad (31a, b, c)$$

We substitute (28) to (30) in (31) and obtain three simultaneous homogeneous equations for B_1 , B_2 , and B_3 . The possible frequencies follow from the determinantal equation

$$\begin{vmatrix} J_0(\mu_1 a)/\mu_1 & J_0(\mu_2 a)/\mu_2 & J_0(\mu_3 a)/\mu_3 \\ J_1(\mu_1 a) & J_1(\mu_2 a) & J_1(\mu_3 a) \\ J_1(\mu_1 a) & J_1(\mu_2 a) & J_1(\mu_3 a) \\ \hline (\alpha a)^2 - (\mu_1 a)^2 & (\alpha a)^2 - (\mu_2 a)^2 & (\alpha a)^2 - (\mu_3 a)^2 \end{vmatrix} = 0. \quad (32)$$

In Eq. (32) we must introduce the values of μ_1 , μ_2 , and μ_3 which follow from the cubic (25). Comparison of (32) and (26) shows that the frequencies of free vibrations are determined by a formula of the form

$$\omega = \frac{t}{a^2} \left[\frac{E}{(1-\nu^2)\rho} \right]^{\frac{1}{2}} F\left(\frac{h}{t}, \frac{h}{a}, \nu\right). \quad (33)$$

For a flat plate ($h=0$) Eq. (33) becomes, for flexural vibrations, $\omega = (t/a^2)[E/(1-\nu^2)\rho]^{\frac{1}{2}} \times \text{const.}$, and for extensional vibrations, $\omega = (1/a)[E/(1-\nu^2)\rho]^{\frac{1}{2}} \times \text{const.}$

The simultaneous solution of (32) and (25), by numerical methods, is presumably not an easy matter and is not given in this note. Instead, a simple approximate formula is derived in what follows by means of the Rayleigh-Ritz procedure.

6. APPROXIMATE DETERMINATION OF LOWEST FREQUENCY BY VARIATIONAL METHOD

The variational form of the equations of motion is according to (13) and (14)

$$\int_0^a C[L(ru) + m_1^2 rw' + \alpha^2 ru] \delta u + D[\{L(rw')\}' + m_2^2 (ru)' - \beta^2 rw] \delta w \, dr = 0. \quad (34)$$

To obtain an approximate value of the lowest frequency we set in (34)

$$\delta u = (\delta A) \frac{r}{a} \left(1 - \frac{r}{a} \right), \quad (35a)$$

and

$$\delta w = (\delta B)(1 - r^2/a^2)^2, \quad (35b)$$

which leads to the following two simultaneous equations for A and B :

$$\left[\frac{1}{60}(\alpha a)^2 - \frac{1}{4} \right] A - \frac{11}{105} m_1^2 a B = 0 \quad (36a)$$

and

$$\frac{11}{105} m_2^2 a^3 A - \left[\frac{1}{10}(\beta a)^4 - \frac{32}{5} \right] B = 0. \quad (36b)$$

The condition of vanishing determinant of (36) may be written in the form

$$\lambda^2 - \left[\frac{80}{9} + 8(1+\nu) \frac{h^2}{t^2} + 15 \frac{a^2}{t^2} \right] \lambda + \frac{a^2}{t^2} \left[\frac{400}{3} + 120(1+\nu) \frac{h^2}{t^2} - 96 \left(\frac{11}{21} \right)^2 (1+\nu)^2 \frac{h^2}{t^2} \right] = 0, \quad (37)$$

where

$$\lambda = (1-\nu^2) \omega^2 \rho a^4 / E t^2. \quad (38)$$

As $(t/a) \ll 1$ and $(h/a) \ll 1$, the smaller of the two roots of (37) simplifies to

$$\lambda = \frac{80}{9} \left[1 + \left(\frac{9}{10} - \frac{242}{1225} (1+\nu) \right) (1+\nu) \frac{h^2}{t^2} \right], \quad (39)$$

and the lowest characteristic frequency is then approximated by the simple formula

$$\omega = 2.98 \frac{t}{a^2} \left(\frac{E}{\rho(1-\nu^2)} \right)^{\frac{1}{2}} \left(1 + (1+\nu) \left[.90 - .20(1+\nu) \right] \frac{h^2}{t^2} \right)^{\frac{1}{2}}. \quad (40)$$

In the limiting case of a flat plate Eq. (40) gives a value which lies within about one percent of the exact value.

It may be noted that had one used, instead of (35), the still rougher approximation $u=0$ and $w=B(1-r^2/a^2)^2$, then instead of (40), an approximate value for ω would have been obtained where the term $.90 - .20(1+\nu)$ is replaced by $.90$.³ This indicates that the upper limit for ω , which is given by (40), should be improved, especially for appreciable values of h/t , either by solving the determinantal equation (32) or by using additional terms in the Rayleigh-Ritz procedure.

³ By using $u = (r/a)(1-r^2/a^2)$ instead of (35a), Federhofer obtains a result which may be transformed into an expression of the form (40), with the difference that the factor $.20$ is replaced by a factor $.19$.

Non-Uniform Transmission Lines and Reflection Coefficients

L. R. WALKER AND N. WAX

Bell Telephone Laboratories, Inc., New York, New York

(Received July 30, 1946)

A first-order differential equation for the voltage reflection coefficient of a non-uniform line is obtained and it is shown how this equation may be used to calculate the resonant wave-lengths of tapered lines.

THE usual method of treating non-uniform (or tapered) transmission line problems consists in solving a second-order differential equation in either the voltage or current along the line. Pierce¹ has shown, recently, how one may obtain a first-order differential equation for the input impedance of a non-uniform line. It is the purpose of this note to derive a first-order differential equation for the voltage reflection coefficient of a non-uniform line and to show how this equation may be used to calculate the resonant wave-lengths of tapered lines.

The equations satisfied by the voltage, V , and current, I , at a point x along any transmission line are

$$dV/dx = -Z(x)I, \quad (1)$$

$$dI/dx = -Y(x)V, \quad (2)$$

where $Z(x)$ and $Y(x)$ are the series impedance and shunt admittance per unit length of line, respectively. The nominal characteristic impedance, $K(x)$, and nominal propagation factor $\Gamma(x)$, are defined by the relations

$$K(x) = \left[\frac{Z(x)}{Y(x)} \right]^{1/2}, \quad (3)$$

$$\Gamma(x) = [Z(x)Y(x)]^{1/2}. \quad (4)$$

The voltage reflection coefficient, q_v , is defined by

$$q_v = \frac{\frac{V}{I} - K(x)}{\frac{V}{I} + K(x)}. \quad (5)$$

Making use of the above equations one finds that q_v satisfies the first-order non-linear differential equation

¹ J. R. Pierce "Note on the transmission line equations in terms of impedance," *Bell Sys. Tech. J.* 22, 263-265 (1943).

$$\frac{dq_v}{dx} - 2\Gamma(x)q_v + \frac{1}{2} \frac{d \log K(x)}{dx} (1 - q_v^2) = 0. \quad (6)$$

Introducing the complex phase angle, θ , by means of the relation $q_v = e^{i\theta}$, Eq. (6) becomes

$$\frac{d\theta}{dx} = -2i\Gamma(x) + \frac{d \log K(x)}{dx} \sin \theta \quad (7)$$

or, in integral form,

$$\theta = -2i \int^x \Gamma(x) dx + \int^x \sin \theta \frac{d}{dx} [\log K(x)] dx. \quad (8)$$

The first term on the right-hand side of Eq. (8) represents a generalization of the usual expression for the phase shift along a uniform line that takes into account the variations in propagation factor. The second term is in the nature of a correction term which depends not only on the changes in nominal characteristic impedance but also on the phase at the point.

Thus far no assumptions have been made concerning the properties of the transmission line. If one now assumes, as will be done in the rest of this paper, that the line is loss-less, then the nominal characteristic impedance will be real and the propagation factor will be purely imaginary, $\Gamma(x) = i\beta(x)$. A new variable, y , may be defined by

$$y = \int_0^x \beta(x) dx.$$

Equations (7) and (8) then become

$$\frac{d\theta}{dy} = 2 + \frac{d \log K(x)}{dy} \sin \theta, \quad (9)$$

$$\theta = \theta(0) + 2y + \int_0^y \sin \theta \frac{d}{dx} [\log K(x)] dx, \quad (10)$$

where $\theta = \theta(0)$ at $x = 0$.

It may be instructive to consider Eq. (7) from the viewpoint of transducer theory. The relation between the reflection coefficients q_{v1} and q_{v2} (looking in the same direction) on the two sides of a loss-less, linear, passive, four-terminal network is

$$q_{v1} = \exp(-i\alpha_1) \frac{\gamma + \exp(i\alpha_2)q_{v2}}{1 + \gamma \exp(i\alpha_2)q_{v2}}, \quad (11)$$

where $\alpha_1, \alpha_2, \gamma$ are all real, and $|\gamma| \leq 1$. Equation (11) may be obtained in a variety of ways, for example from the definition of the voltage reflection coefficient and the known relationships between the voltages and currents on the two sides of the network. Let $q_{v1} = \exp(i\theta_1)$ and $q_{v2} = \exp(i\theta_2)$. Then Eq. (11) becomes

$$\tan\left(\frac{\theta_2 + \alpha_2}{2}\right) = \Delta \tan\left(\frac{\theta_1 + \alpha_1}{2}\right), \quad (12)$$

where $\Delta = (1 + \gamma)/(1 - \gamma)$. If one now considers a small element of the transmission line as an elementary four-terminal network, then $\theta_2 = \theta_1 + \delta\theta$, $\alpha_1 = \delta\alpha_1$, $\alpha_2 = \delta\alpha_2$, and $\Delta = 1 + \delta\Delta$. One finds that

$$\delta\theta = (\delta\alpha_1 - \delta\alpha_2) + \delta\Delta \sin \theta_1, \quad (13)$$

and comparing this with Eq. (7)

$$\delta\alpha_1 - \delta\alpha_2 = \beta(x)\delta x, \quad (14)$$

$$1 + \delta\Delta = 1 + \delta \log K(x) = \frac{K(x) + \delta K(x)}{K(x)}. \quad (15)$$

An equivalent representation for each small section of a non-uniform line is, therefore, two small lengths of uniform line with the characteristic impedance of one line that of the nominal characteristic impedance at the beginning of the section, and that of the other the characteristic impedance at the end of the section.

Equation (10) may be solved by the method of iteration; the approximations converge rapidly when the change in $\log K(x)$ over the interval considered is sufficiently small. The n th-order approximation will be

$$\theta_n = \theta(0) + 2\gamma + \int_0^x \sin \theta_{n-1} \frac{d}{dx} [\log K(x)] dx. \quad (16)$$

The zero-order approximation is given by

$$\theta_0 = \theta(0) + 2\gamma \quad (17)$$

and the first-order approximation by

$$\theta_1 = \theta(0) + 2\gamma + \int_0^x \sin \theta_0 \frac{d}{dx} [\log K(x)] dx. \quad (18)$$

Equations (16)–(18) enable one to solve Eq. (10) when the propagation factor is known. The calculation of the resonant wave-lengths of non-uniform lines presents a somewhat different problem in that the propagation factor is the quantity to be determined.

It is convenient to make two more assumptions now. One is that the line is terminated in a pure reactance at $x=0$; this makes θ a real quantity. The other is that the propagation factor is a constant as a function of x ; an important special class of non-uniform lines satisfies this condition. Equation (10) becomes, then,

$$\theta = \theta(0) + \frac{4\pi}{\lambda}x + \int_0^x \sin \theta \frac{d}{dx} [\log K(x)] dx, \quad (19)$$

where $\beta(x) = \beta$; $\beta = 2\pi/\lambda$.

Resonance will occur when $\theta = m\pi$, where m is an even positive integer, anti-resonance when m is an odd positive integer.

Let $x=a$ when $\theta = m\pi$; in solving Eq. (19) by the method of iteration one sets all approximations equal to $m\pi$ at $x=a$.

The n th-order approximation to λ is given by

$$\lambda_n = \frac{4a}{m - \frac{\theta(0)}{\pi} - \frac{1}{\pi} \int_0^a \sin \theta_{n-1} \frac{d}{dx} [\log K(x)] dx} \quad (20)$$

with the zero-order approximation

$$\lambda_0 = \frac{4a}{\theta(0) - \frac{m\pi}{\pi}} \quad (21)$$

and the first-order approximation

$$\lambda_1 = \frac{4a}{m - \frac{\theta(0)}{\pi} - \frac{1}{\pi} \int_0^a \sin \theta_0 \frac{d}{dx} [\log K(x)] dx} \quad (22)$$

Resonant wave-lengths of a few non-uniform lines have been obtained by the standard meth-

ods and by means of Eq. (22); the results were in agreement when the taper was small. The appendix gives the results of such a calculation for a particular class of non-uniform lines, together with the errors arising from the use of the first approximation.

APPENDIX

If one assumes that the propagation factor is constant, and that $K(x)$ varies as a power of x , then Eq. (9) will be of the form

$$\frac{d\theta}{dx} = 2\beta - \frac{1 \pm 2\nu}{x} \sin \theta \quad (\text{A1})$$

with β and ν real constants.

A new variable w may be introduced by means of the transformation

$$\frac{1}{w} \frac{dw}{dx} = \pm \frac{\nu}{x} - \beta \tan \theta/2. \quad (\text{A2})$$

Using (A1) and (A2) one obtains Bessel's equation

$$\frac{d^2w}{dx^2} + \frac{1}{x} \frac{dw}{dx} + \left(\beta^2 - \frac{\nu^2}{x^2} \right) w = 0, \quad (\text{A3})$$

with solutions, therefore, of the form

$$w = A J_\nu(\beta x) + B Y_\nu(\beta x) \quad (\text{A4})$$

A and B are constants.

If $w=0$ at $x=a$ and at $x=b$ then (A4) yields

$$J_\nu(\beta b) Y_\nu(\beta a) = J_\nu(\beta a) Y_\nu(\beta b). \quad (\text{A5})$$

From (A2), the zeros of w are the poles of $\tan \theta/2$, for $x \neq 0$; the corresponding boundary conditions on (A1) are then $\theta = (2m+1)\pi$ at $x=a$, $\theta = (2n+1)\pi$ at $x=b$, where m and n are integers. The eigenvalues of (A1), subject to these boundary conditions, are thus the roots of (A5). One may readily obtain a first approximation to the eigenvalues of (A1), and therefore to the roots of (A5) by using the method of iteration. This

yields

$$\beta_1 = \frac{1}{b-a} \left\{ s\pi - \frac{1 \pm 2\nu}{2} \times \left[\left(Si\left(\frac{2s\pi a}{\alpha-1}\right) - Si\left(\frac{2s\pi}{\alpha-1}\right) \right) \cos \frac{2s\pi}{\alpha-1} + \left(Ci\left(\frac{2s\pi a}{\alpha-1}\right) - Ci\left(\frac{2s\pi}{\alpha-1}\right) \right) \sin \frac{2s\pi}{\alpha-1} \right] \right\}, \quad (\text{A6})$$

$s=m-n$, $\alpha=b/a > 1$, Si and Ci are the sine- and cosine-integrals, respectively.

Roots of (A5) may be obtained quite quickly by means of (A6) and are of fair accuracy when ν and α are small. The percentage error for $\nu=0$ and $\alpha \leq 4$ is less than 2.5 percent for the first five roots; for $\nu=\frac{1}{2}$ and $\alpha \leq 4$ it is less than 6.2 percent for the first three roots; for $\nu=1$ and $\alpha \leq 2.5$ it is less than 5.5 percent for the first four roots. When $\nu=2$ and $\alpha=2$, the first root is in error by 13.2 percent. In all cases computed the percentage error decreased rapidly as s increased.

The above approximation has been compared with the McMahan² series solution for (A5). It was found that for $\alpha < 4$, $\nu < 2$, a single root could be obtained with about equal speed by the two methods, for comparable accuracy. However, the above approximation has two advantages over the McMahan series: (1) that approximate roots may be obtained very quickly for many values of α , and (2) that even a single root may be obtained with greater speed for $\alpha > 4$, since the McMahan series converges slowly for these values.

ACKNOWLEDGMENTS

The authors wish to thank their colleagues at the Bell Telephone Laboratories and in particular Dr. S. A. Schelkunoff for his help during the preparation of this paper, and Miss M. E. Moore for computing aid.

² A. Gray, G. B. Mathews, and T. M. MacRobert, *A Treatise on Bessel Functions* (The Macmillan Company, New York, 1922), Appendix III (VI), p. 261.

Resonant Frequencies of the Nosed-In Cavity

ERNEST MAYER*

Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania

(Received June 13, 1946)

In this paper the resonant frequency of the nosed-in cavity is studied as a function of the cavity dimensions. Maxwell's differential equations and boundary conditions are converted into an integral equation which is solved approximately by the Ritz variational method. The size and shape of the cavity are fixed by specification of the dimensions (cf. Fig. 1) r_1 and r_2 , the inner and outer radii; ϵ^I , the post length and ϵ^{II} , the gap space. If the cavity is to resonate to the wave-length λ , only three of its dimensions can be given independently; the fourth will be a function of the given three and the wave number $k = 2\pi/\lambda$. For fixed r_1/r_2 the dependence of $k\epsilon^I$ on $k\epsilon^{II}$ is calculated with a precision of 1 percent.

I. INTRODUCTION

THE present article is based almost entirely on a Westinghouse Research report.¹ A number of mathematical details and proofs not given here and additional references will be found in the above-mentioned report.

Much interest has developed recently in the nosed-in type of cavity of large gap spacing shown in Fig. 1. In this article we obtain curves which relate the resonant frequency of the cavity to its dimensions. These dimensions are: ϵ^I , the post length; ϵ^{II} , the gap space; r_1 , the post radius; and r_2 , the outer radius of the cavity. As the cavity is a figure of revolution we use cylindrical polar coordinates (r, ϕ, z) with the z axis along the axis of symmetry and the $z=0$ plane along the tip of the post. We are interested in the fundamental mode of resonance. In this mode the angular component of the electric field, E_ϕ , is zero and the angular component of the magnetic field, H_ϕ , is independent of ϕ . Since the mode has angular symmetry, H_ϕ satisfies the partial differential equation

$$\frac{\partial^2 H_\phi}{\partial r^2} + \frac{1}{r} \frac{\partial H_\phi}{\partial r} + \frac{\partial^2 H_\phi}{\partial z^2} + \left(k^2 - \frac{1}{r^2}\right) H_\phi = 0, \quad (1)$$

together with the boundary condition that, on all conducting surfaces,

$$\frac{\partial}{\partial n}(rH_\phi) = 0, \quad (2)$$

where $\partial/\partial n$ denotes differentiation in the direction normal to the conducting surface, and $k = 2\pi/\lambda$ is the wave number. H_ϕ can serve as a potential function for the other two non-vanishing components of the field:²

$$ikE_z = -\frac{1}{r} \frac{\partial}{\partial r}(rH_\phi), \quad (3)$$

$$ikE_r = -\frac{\partial H_\phi}{\partial z}. \quad (4)$$

For the cavity in Fig. 1 no explicit solution of (1) has been given. Only approximate methods are available for calculating the frequency of such a resonator. A practical procedure is to divide the interior of the cavity by a suitably chosen surface S into regions I and II, in each of which the variables of (1) can be separated. In these regions the solutions H_ϕ^I and H_ϕ^{II} , respectively, can be set up so as to satisfy (2) at the conducting surfaces. However, owing to the fact that neither region is completely surrounded by conducting walls, each solution will contain an infinite number of arbitrary constants; furthermore, the wave

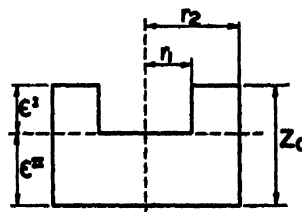


FIG. 1.

* Now at The M. W. Kellogg Company, Jersey City, New Jersey.

¹ T. Holstein and E. Mayer, Research Report SR-281, Resonant frequencies of the nosed-in cavity.

² E. U. Condon, Rev. Mod. Phys. 14, 341 (1942).

number k is undetermined. In order for H_ϕ^I and H_ϕ^{II} to represent the solution of (1) in the whole cavity, they must obey the following continuity conditions on the surface S :

$$H_\phi^I = H_\phi^{II}, \quad (5)$$

$$\partial H_\phi^I / \partial n = \partial H_\phi^{II} / \partial n, \quad (6)$$

where $\partial/\partial n$ denotes differentiation in the direction normal to S . The imposition of these continuity conditions determines not only the arbitrary constants but also the wave number k .

This procedure has been applied by Hansen³ to the nosed-in resonator of small gap space. In his paper the surface S is chosen at $r=r_1$. The interior is thereby divided into two regions in each of which H_ϕ can be expanded as a Fourier-Bessel summation, the amplitudes of the terms being determined by the condition that both summations yield the same value of the tangential electric field E_s on the dividing surface S . This condition, which is equivalent to (6), together with (5) can be used to determine the unknown amplitudes and the wave number k .

Hansen now assumes an analytic form $E_s(z)$ to approximate the component of the actual field on S . The assumed $E_s(z)$ is expanded into two Fourier series over the intervals ϵ^I and $\epsilon^I + \epsilon^{II} = z_0$, corresponding to the regions into which Hansen's S divides the interior of the cavity. Substitution of the Fourier expansions into (3) permits evaluation of the amplitudes and therefore satisfies (6) by construction. The determination of the only remaining unknown, k , is now effected by a matching of H_ϕ^I and H_ϕ^{II} at z_1 , a particular value of z on the surface S employed by Hansen. With the value k thus found, say k_1 , the matching indicated in (5) is in general not satisfied at $z \neq z_1$, because the analytic form of $E_s(z)$ is only an approximation. Had E_s been the correct function, the k obtained from matching at a particular z would have guaranteed satisfaction of (5) everywhere on S . Hansen shows that k_1 is close to the correct k as it is not very sensitive to the error in E_s or to the choice of z_1 . In practical calculations it turns out that his method begins to be inaccurate for resonators whose total height z_0 exceeds $\pi/2k$.

A more general approach to the problem has

³ W. W. Hansen, J. App. Phys. 10, 1 (1939).

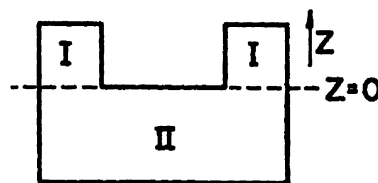


FIG. 2.

been made by J. Schwinger. According to his method E_s (or, more generally, the tangential electric field on S) is left undetermined. The amplitudes of the Fourier terms are expressed by means of (3) and (6) as integrals of the unknown E_s . Substitution in (5) leads to an integral equation in E_s of which solutions exist only for a denumerably discrete set of k 's. In applying such integral equations to wave guide problems Schwinger has shown how approximate solutions can be obtained by variational methods.

The frequency calculations presented in this paper extend beyond the range treated by Hansen, i.e., the method presented here is practical for $z_0 > \pi/2k$. These calculations are based on an integral equation of the form

$$\int \psi(r) G(r, r') dr' = 0, \quad (7)$$

in which the kernel $G(r, r')$ is a function of the cavity dimensions and the wave number k . The steps leading to (7) are stated below. The solution of the integral equation will be achieved by methods analogous to those employed by Schwinger.

II. THE INTEGRAL EQUATION

The derivation of the integral equation proceeds as follows. We divide the cavity interior into regions I and II by choosing* S along $z=0$ as shown in Fig. 2. The solutions of (1) in both regions are of the form

$$H_\phi = Z_1(k, r) e^{ikz},$$

where

$$k_r^2 + k_z^2 = k^2$$

and $Z_1(k, r)$, the first-order cylindrical function, is a linear superposition of Bessel and Neumann

* This choice of S is merely a matter of mathematical convenience. The variational method applied here can be applied also when S is chosen at $r=r_1$ as in Hansen's paper (see reference 3).

functions

$$Z_1(k, r) = aJ_1(k, r) + bN_1(k, r),$$

a and b being constants of integrations. Using H_ϕ as a potential from which E_s and E_r can be derived according to (3) and (4), we can now impose the boundary condition that the tangential electric field shall vanish on all conducting surfaces, and thereby arrive at relations between a and b and between k_s and k from which simplified expressions for II_ϕ^I and II_ϕ^{II} can be written in closed form.

We start with region II, in which $b=0$ since the Neumann function is excluded because of its singularity at $r=0$. Further, the boundary condition $E_s=0$ at $r=r_2$ leads to $J_0(k, r_2)=0$. Hence, k, r_2 can have any of the values j_n/r_2 , where j_n is the n 'th root of the zero-order Bessel function. In cavities of practical interest $kr_2 < j_1 = 2.405$, and, therefore, all the k, r_2 's are imaginary with

$$-(k, r_2)^2 = (j_n/r_2)^2 - k^2 = (k_n^{II})^2, \quad k_n^{II} \text{ real.} \quad (8)$$

We can thus express H_ϕ^{II} as

$$H_\phi^{II} = \sum A_n^{II} \cosh k_n^{II}(z + \epsilon^{II}) J_1(j_n r/r_2). \quad (9)$$

Here the phase of the cosh term has been adjusted to make E_r zero on the wall at $z = -\epsilon^{II}$.

In region I we find that in order for E_s to vanish at $r=r_1$ and $r=r_2$ we must have

$$Z_0(k, r_1) = 0 \quad \text{and} \quad Z_0(k, r_2) = 0.$$

From the latter two conditions we can, in the first place, solve for a/b so that we can write Z_1 as

$$Z_1(k, r) = -N_0(k, r_1) J_1(k, r) + J_0(k, r_1) N_1(k, r) \quad (10)$$

and secondly we find that k, r_1 must satisfy

$$J_0(k, r_1) N_0(k, r_2) - N_0(k, r_1) J_0(k, r_2) = 0. \quad (11)$$

Introducing the notation $\zeta = k, r_2$ we rewrite (11):

$$J_0(\zeta r_1/r_2) N_0(\zeta) - N_0(\zeta r_1/r_2) J_0(\zeta) = 0. \quad (11')$$

The solutions $\zeta = \zeta_n$ of this equation are listed in Jahnke-Emde's⁴ *Tables of Functions*. As ζ_n is always larger than the corresponding j_n , we again have imaginary k, r_2 's or

$$-(k, r_2)^2 = (\zeta_n/r_2)^2 - k^2 = (k_n^I)^2, \quad k_n^I \text{ real.} \quad (12)$$

In addition to the Z_1 's listed in (10) the function $1/r$ is a solution for the case $k, r=0$ since it satisfies (1) and its boundary conditions. Thus the complete solution of (1) in region I is

$$II_\phi^I = A_0^I \cos k(z - \epsilon^I)/r + \sum A_n^I \cosh k_n^I(z - \epsilon^I) Z_1(\zeta_n r/r_2), \quad (13)$$

where the dependence on z has been so adjusted that E_r vanishes on the wall at $z = \epsilon^I$.

We now define $\psi(r)$ by the relation

$$\psi(r)/r = \partial II_\phi / \partial z = -ikE_r \quad \text{at} \quad z=0. \quad (14)$$

This is the ψ which appears in (7) as the unknown function. In solving (7) for the wave number k we shall obtain, as a by-product of the calculations, the dependence of ψ on r and hence the functional form of E_r .

Following the procedure outlined above we solve for the amplitude A_n as integrals of the tangential field (in this case E_r) on S . Substitution of (9) and (13) in (14) gives

$$\frac{\psi(r)}{r} = \sum A_n^{II} k_n^{II} \epsilon^{II} J_1(j_n r/r_2), \quad (15)$$

$$\frac{\psi(r)}{r} = A_0^I k \sin k \epsilon^I / r - \sum A_n^I k_n^I \sinh k_n^I \epsilon^I Z_1(\zeta_n r/r_2). \quad (16)$$

To solve for A_n^{II} we multiply in (15) by $J_1(j_m r/r_2) r dr$ and integrate over S obtaining

$$A_n^{II} = \frac{2k_n^{II} \int_{r_1}^{r_2} J_1(j_m r'/r_2) \psi(r') dr'}{(k_n^{II} r_2)^2 \sinh k_n^{II} \epsilon^{II} J_1^2(j_m)}. \quad (17)$$

⁴E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943), p. 205. The tables from which ζ_n can be obtained contain a number of errors. New and more accurate values were taken from tables by A. N. Lowan and A. Hillman, *J. Math. Phys.* 22 (Dec., 1943). These tables contain suitable information in the range $r_2/r_1 \leq 4$. For $r_2/r_1 > 4$ we calculated ζ_n 's from asymptotic expansions quoted by Lowan and Hillman.

An analogous integration of (16) with $Z_1(\xi_m r/r_2) r dr$ leads to

$$A_m^I = \frac{-2k_m^I \int_{r_1}^{r_2} Z_1(\xi_m r'/r_2) \psi(r') dr'}{(k_m^I r_2)^2 \sinh k_m^I \epsilon^I [Z_1^2(\xi_m) - (r_1/r_2)^2 Z_1^2(\xi_m r/r_2)]}. \quad (18)$$

The Z_1 's in the denominator of (18) can be expressed in terms of J_0 's⁵ so that A_m^I becomes

$$A_m^I = -\frac{\frac{\pi^2}{2} k_m^I \xi_m^2 \int_{r_1}^{r_2} Z_1(\xi_m r'/r_2) \psi(r') dr'}{(k_m^I r_2)^2 \sinh k_m^I \epsilon^I [(J_0(\xi_m r_1/r_2)/J_0(\xi_m))^2 - 1]}. \quad (18')$$

Finally to obtain A_0^I we integrate (16) with $1/r \cdot r dr$ and get

$$A_0^I = \int_{r_1}^{r_2} \frac{\psi(r')}{r'} dr' / k \sin k \epsilon^I \log \frac{r_2}{r_1}. \quad (19)$$

Having obtained the expressions for A_n we can formulate the boundary condition (5) as

$$H_\phi^I - H_\phi^{II} = 0 = \int_{r_1}^{r_2} \psi(r') G(r, r') dr', \quad (7')$$

where

$$G(r, r') = \frac{\cot k \epsilon^I}{k \log r_2/r_1} \frac{1}{r} \frac{1}{r'} - \frac{\pi^2}{2} \sum \frac{k_n^I \coth k_n^I \epsilon^I Z_1(\xi_n r/r_2) Z_1(\xi_n r'/r_2)}{(k_n^I r_2)^2 [(J_0(\xi_n r_1/r_2)/J_0(\xi_n))^2 - 1]} - 2 \sum \frac{k_n^{II} \coth k_n^{II} \epsilon^{II}}{(k_n^{II} r_2)^2 J_1^2(j_n)} J_1(j_n r/r_2) J_1(j_n r'/r_2). \quad (20)$$

III. THE EQUIVALENT VARIATIONAL PROBLEM

The solution of the integral Eq. (7') is achieved by the solution of the equivalent variational problem, which we shall now proceed to formulate. We define

$$U = \iint \frac{\psi(r) G'(r, r') \psi(r') dr dr'}{\frac{\cot k \epsilon^I}{k \log r_2/r_1} \iint \frac{\psi(r)}{r} \frac{\psi(r')}{r'} dr dr'}, \quad (21)$$

where

$$G'(r, r') = \frac{\cot k \epsilon^I}{k \log r_2/r_1} \frac{1}{r} \frac{1}{r'} - G(r, r')$$

and all the integrations extend, as in the subsequent expressions where the limits are omitted, from r_1 to r_2 . The two equations

$$U = 1 \quad (22) \quad \text{and} \quad \delta U = 0, \quad (23),$$

where δU is the variation in U arising from an arbitrary variation $\delta\psi$ of ψ , then constitute a variational problem equivalent to (7').

⁵ G. N. Watson, *Theory of Bessel Functions* (Cambridge, 1922), p. 76.

To prove this equivalence we perform the variation

$$\delta U = \frac{2 \int \delta \psi(r) dr}{\int \int \frac{\cot k\epsilon^I}{k \log r_2/r_1} \frac{\psi(r)}{r} \frac{\psi(r')}{r'} dr dr'} \left[\int G'(r, r') \psi(r') dr' - \frac{\int \int \psi(r) G'(r, r') \psi(r') dr dr' \int \frac{\cot k\epsilon^I}{k \log r_2/r_1} \frac{\psi(r')}{r r'} dr'}{\int \int \frac{\cot k\epsilon^I}{k \log r_2/r_1} \frac{\psi(r)}{r} \frac{\psi(r')}{r'} dr dr'} \right]$$

Since $\delta\psi$ is arbitrary, the bracketed quantity must vanish; or, in view of (21)

$$\int G'(r, r') \psi(r') dr' - U \int \frac{\cot k\epsilon^I}{k \log r_2/r_1} \frac{\psi(r')}{r r'} dr' = 0 \quad (23')$$

which, for $U=1$, is identical with (7').

IV. PROPERTIES OF THE SOLUTION

The theory of integral equations asserts that the solution of (23) (or its equivalent (23')) is in general an infinite set of eigenfunctions ψ_n which belong to corresponding terms of a set of eigenvalues U_n of U . Owing to the particular form of $G'(r, r')$ a number of special features of the solution, outlined below in (a), (b), and (c), can be proved.¹ These features serve to guide us in the application of approximate methods for solving (23).

For specified values of ϵ^I , ϵ^{II} , r_1 , r_2 , and k in the range

$$kr_2 < j_1 = 2.405, \quad k\epsilon^I < \pi/2, \quad (24)$$

the following theorems hold:

(a) There exist only one eigenvalue (denoted by U_0) and one corresponding eigenfunction (denoted by ψ_0) which satisfy (23).

(b) Equation (22) with $U=U_0$, that is $U_0=1$, constitutes an equation among the five parameters from which ϵ^I can be evaluated uniquely when the other four are specified. (The value of ϵ^I so obtained will be denoted by ϵ_0^I .)

(c) The value of ϵ^I obtained from (22) with an arbitrary ψ cannot exceed ϵ_0^I ; i.e., $\epsilon^I \leq \epsilon_0^I$.

In view of (c) we may expect that² if we find

¹ R. Courant and D. Hilbert, *Methoden der Mathematischen Physik* (Interscience Publishers, Inc., New York, 1943), Vol. 1, p. 146.

□

a sequence of ψ 's such that the corresponding ϵ^I 's form a monotonically increasing series, the limit of this series is ϵ_0^I .

V. THE RITZ VARIATIONAL METHOD

In this section we describe the Ritz procedure by which an approximate solution of (22) and (23) can be obtained. We select n functions $\psi_j(r)$ each of which can be made to satisfy the boundary conditions on $\psi(r)$ and set up

$$\psi(r) = \sum_{j=1}^n \alpha_j \psi_j(r). \quad (25)$$

Variations in ψ are now effected by variations in α_j . With (25) substituted in (22) and (23) we can determine α_j and solve for ϵ^I as a function of ϵ^{II} , r_1 , r_2 and k .

In using (25) we find the following notation convenient:

$$G_{ji} = \int \int \psi_j(r) G'(r, r') \psi_i(r') dr dr' = G_{ji},$$

$$I_{ji} = \frac{\cot k\epsilon^I}{k \log (r_2/r_1)} \int \int \frac{\psi_j(r)}{r} \frac{\psi_i(r')}{r'} dr dr' = I_{ji}.$$

With these formulas we convert (22) and (23) into

$$\frac{\sum_i \sum_j \alpha_i \alpha_j G_{ji}}{\sum_i \sum_j \alpha_i \alpha_j I_{ji}} = 1$$

and

$$\delta \frac{\sum_i \sum_j \alpha_i \alpha_j G_{ij}}{\sum_i \sum_j \alpha_i \alpha_j I_{ij}} = 0, \quad i, j = 1, 2 \dots n,$$

respectively. The variation is performed on α ; giving

$$\frac{2 \sum_i \delta \alpha_i}{\sum_i \sum_j \alpha_i \alpha_j I_{ij}} \left[\sum_i \alpha_i G_{ii} - \sum_i \alpha_i I_{ii} \frac{\sum_i \sum_j \alpha_i \alpha_j G_{ij}}{\sum_i \sum_j \alpha_i \alpha_j I_{ij}} \right] = 0.$$

As the $\delta \alpha_i$'s are arbitrary, the bracketed term must vanish; or, after simplification,

$$\sum_{i=1}^n \alpha_i (G_{ii} - I_{ii}) = 0; \quad j = 1, 2 \dots n.$$

This system of n linear equations for the $n \alpha_i$'s has a non-trivial solution if, and only if, the determinant formed from the coefficients of the α_i 's vanishes;

$$\begin{vmatrix} G_{11} - I_{11} & G_{12} - I_{12} & \dots \\ G_{21} - I_{21} & G_{22} - I_{22} & \dots \\ \dots & \dots & \dots \end{vmatrix} \equiv 0. \quad (26)$$

This expression constitutes a transcendental equation from which it is possible to solve for ϵ^1 when the other four parameters are specified.

VI. THE CHOICE OF Ψ 'S

In choosing suitable ψ_i 's, we take into account the behavior of the correct solution ψ_0 and its derivatives on S . Thus as $r \rightarrow r_2$, we must have $E_r \rightarrow 0$ and hence from (14) and

$$\text{div } E = \frac{\partial}{\partial r} E_r + r \frac{\partial E_r}{\partial z} = 0,$$

it follows that

$$\partial \psi_0 / \partial r = 0 \quad \text{at } r = r_2. \quad (27a)$$

When $r \rightarrow r_1$, $E_r \rightarrow \infty$ as $(r - r_1)^{-1/2}$ so that we have

$$\psi_0(r) \propto (r - r_1)^{-1/2}, \quad (27b)$$

$$\partial \psi_0 / \partial r \propto (r - r_1)^{-3/2}, \quad \text{as } r \rightarrow r_1.$$

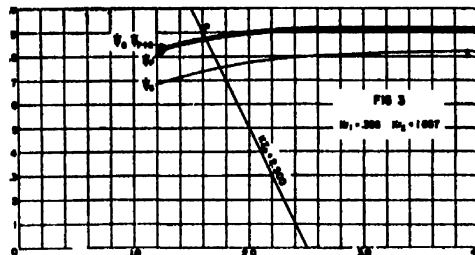


FIG. 3.

Finally, from the analogous two-dimensional case it is to be expected that

$$\partial^2 \psi_0 / \partial r^2 > 0 \quad \text{on } S. \quad (27c)$$

Now, we seek a set of ψ_i 's which can be combined according to (25) to make ψ behave approximately as ψ_0 . A vital property of our ψ_i 's must be, however, that they permit the evaluation of the integrals for G_{ij} and I_{ij} . The question of this integrability is important enough to warrant the relaxation of some of the conditions in (27) in order that a practical set of ψ_i 's be found.

After a number of trials it was found that a family of ψ_i 's which can be made to agree satisfactorily with conditions (27) and also permit evaluation of $G_{ij} - I_{ij}$ without undue difficulty consists of power series in r . The basic ψ_n 's considered in this family are

$$\psi_0 = 1, \text{ a constant,}$$

$$\psi_p = \alpha_1 + \alpha_2 [(r - r_2)^2 - a_p], \text{ a parabolic } \psi_n,$$

$$\psi_q = \alpha_1' + \alpha_2' [(r - r_2)^4 - a_q], \text{ a quartic } \psi_n,$$

and

$$\psi_{p+q} = \alpha_1'' + \alpha_2'' [(r - r_2)^2 - a_p] + \alpha_2''' [(r - r_2)^4 - a_q],$$

a superposition of parabolic and quartic ψ_n 's. The constants a_p and a_q are determined from

$$\int \frac{(r - r_2)^2 - a_p}{r} dr = \int \frac{(r - r_2)^4 - a_q}{r} dr = 0,$$

which will insure that all the I_{ij} 's vanish except

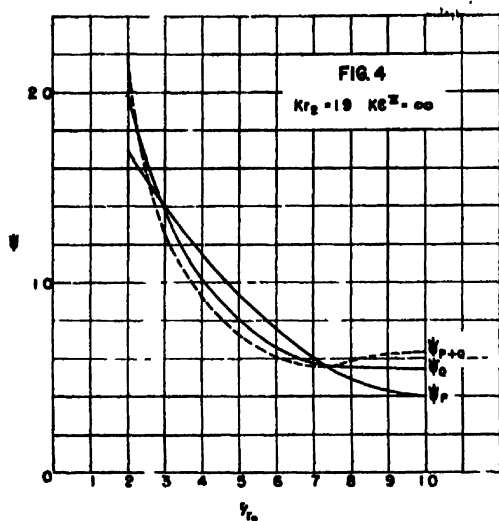


FIG. 4.

I_{11} . For ψ_e the determinantal equation reduces to

$$G_{11} - I_{11} = 0 \quad (26')$$

from which values of $k\epsilon^I$ as function of $k\epsilon^{II}$ with fixed kr_1 and kr_2 have been calculated (Fig. 3). In view of more accurate* final curves, the value of $k\epsilon^I$ thus obtained is about 10 percent below $k\epsilon_0^I$.

Considerable improvement over the ψ_e curve is obtained by use of the parabolic ψ_p , which leads to the determinantal equation

$$\begin{vmatrix} G_{11} - I_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix} = 0. \quad (26'')$$

Further increase in the calculated value of $k\epsilon^I$ is obtained by use of ψ_q . When the more general ψ_{p+q} is employed, there is no appreciable increase in $k\epsilon^I$ over the result with ψ_q , as shown in Fig. 3. The convergence of the Ritz process illustrated in Fig. 3 indicates that ψ_q is sufficiently good for

$$a_q = r_2^4 \frac{12 \log r_2/r_1 - 25 + 48(r_1/r_2) - 36(r_1/r_2)^2 + 16(r_1/r_2)^3 - 3(r_1/r_2)^4}{12 \log r_2/r_1}.$$

From the definitions of I_{11} and G_{ij} we are led to the following formulas for these terms:

$$I_{11} = r_2 \frac{\log r_2/r_1}{kr_2} \cot k\epsilon^I,$$

* Since from (c) in Section IV ϵ_0^I is the upper limit of the ϵ^I that can be calculated from an arbitrary ψ , the highest curve on Fig. 3 is the most accurate.

the calculation of characteristics of the nosed-in resonator, and therefore this function was employed in preference to more general functions, with which the numerical work becomes exceedingly laborious without compensating gain in accuracy. The calculations with ψ_q are given in greater detail in the next section.

The experimental check on these curves was obtained at the Westinghouse Research Laboratories. The resonator of total height $kz_0 = k(\epsilon^I + \epsilon^{II}) = 2.5$ having $k\epsilon^I$ and $k\epsilon^{II}$ corresponding to point P on Fig. 3 shows an agreement within 2.5 percent of the calculated value. Further measurements at the Bloomfield Division of Westinghouse Electric Corporation led to similar agreement, the experimental values of $k\epsilon^I$ lying about 2.5 percent above the calculated curves. The disagreement is ascribed to the deviation of the actual cavity from that treated by the theory. Instead of the smooth top shown in Figs. 1 and 2 the actual cavity is provided with a corrugated diaphragm (to permit tuning).

The radial variation of ψ_p , ψ_q , and ψ_{p+q} is illustrated in Fig. 4. It is to be noted that (27b) is not satisfied by these ψ 's, but this does not result in serious trouble because the infinity of ψ_0 at $r=r_1$ is integrable.

VII. SUMMARY OF PROCEDURE AND RESULTS

A practical set of ψ_n 's consists of the two functions

$$\psi_1 = \psi_e = 1 \quad \text{and} \quad \psi_2 = \psi_q = (r - r_2)^4 - a_q,$$

where the constant a_q , defined by

$$\int_{r_1}^{r_2} \psi_q d\tau / r = 0,$$

is

$$G_{11} = 2r_2 \sum \frac{\coth k_n^{II} \epsilon^{II}}{(k_n^{II} r_2)} \frac{1}{j_n^2} \frac{J_0^2(j_n r_1/r_2)}{J_1^2(j_n)},$$

$$G_{13} = 8r_2^3 \sum \frac{\coth k_n^{II} \epsilon^{II} J_0(j_n r_1/r_2)}{(k_n^{II} r_2) j_n^2 J_1^2(j_n)} E_{3n}^{II}(r_1/r_2) = G_{31},$$

$$G_{33} = 32r_2^3 \left\{ \sum \frac{\coth k_n^I \epsilon^I [E_{3n}^I(r_1/r_2)]^2}{(k_n^I r_2) \zeta_n^2 \left[\left(\frac{J_0(\zeta_n r_1/r_2)}{J_0(\zeta_n)} \right)^2 - 1 \right]} + \sum \frac{\coth k_n^{II} \epsilon^{II} [E_{3n}^{II}(r_1/r_2)]^2}{(k_n^{II} r_2) j_n^2 J_1^2(j_n)} \right\}$$

In these formulas the E 's, which are functions only of the ratio r_1/r_2 are defined by

$$E_{3n}^I = \frac{J_0(\zeta_n r_1/r_2)}{J_0(\zeta_n)} \left[1/\zeta_n - 4/\zeta_n^3 - \frac{\pi}{2} H_0(\zeta_n) (1 - 3/\zeta_n^2) \right]$$

$$- \left[(3 - 3r_1/r_2 + (r_1/r_2)^2)/\zeta_n - 4/\zeta_n^3 - \frac{\pi}{2} H_0(\zeta_n r_1/r_2) (1 - 3/\zeta_n^2) \right],$$

$$E_{3n}^{II} = J_1(j_n) \left[1/j_n - 4/j_n^3 - \frac{\pi}{2} H_0(j_n) (1 - 3/j_n^2) \right]$$

$$- r_1/r_2 J_1(j_n r_1/r_2) \left[(3 - 3r_1/r_2 + (r_1/r_2)^2)/j_n - 4/j_n^3 - \frac{\pi}{2} H_0(j_n r_1/r_2) (1 - 3/j_n^2) \right]$$

$$+ r_1/r_2 J_0(j_n r_1/r_2) \left[\left(-\frac{2}{j_n^2} + \frac{3}{2} \right) (r_1/r_2) - (r_1/r_2)^2 + \frac{1}{2} (r_1/r_2)^3 \right]$$

$$+ \frac{(1 - a_0/r_2^4)}{4r_1/r_2} - \frac{\pi}{2} H_1(j_n r_1/r_2) (1 - 3/j_n^2) \Big],$$

where the H_0 and H_1 are the zero and first-order Struve functions, respectively. Next we set up the determinantal equation

$$\begin{vmatrix} G_{11} - I_{11} & G_{13} \\ G_{31} & G_{33} \end{vmatrix} = 0, \quad (26''')$$

which connects the wave number k and the geometric parameters ϵ^I , ϵ^{II} , r_1 , and r_2 . If, now, k is assigned, this equation can be solved for ϵ^I when the other three parameters are specified provided that $kr_2 < j_1 = 2.405$.

It is convenient to fix the value of r_1/r_2 and let $k\epsilon^{II}$ assume an arbitrary set of values corresponding to which $k\epsilon^I$ is determined by (26'''). Expanding the latter into

$$I_{11} = G_{11} - G_{13}^2/G_{33}, \quad (28)$$

we note that $k\epsilon^I$ appears on the left in I_{11} and $k_n^I \epsilon^I$ on the right in G_{13} , which, however, is not very sensitive to variations in ϵ^I . We can thus first obtain $k\epsilon^I$ approximately from $I_{11} = G_{11}$, calculate from (12)

$$\coth k_n^I \epsilon^I = \coth k\epsilon^I [\zeta_n^2/(kr_2)^2 - 1]^{\frac{1}{2}}$$

and substitute this in G_{13} . It turns out that only $\coth k_1^I \epsilon^I$ need be calculated in this manner because for $n > 1$ $\coth k_n^I \epsilon^I \approx 1$. Similarly, we have from (8) $\coth k_1^{II} \epsilon^{II} = \coth k\epsilon^{II} [j_1^2/(kr_2)^2 - 1]^{\frac{1}{2}}$ and for $n > 1$ $\coth k_n^{II} \epsilon^{II} \approx 1$.^{*} With these expressions for the coth's the summations in the right member of (28) are readily performed up to a

^{*} In order for this approximation to hold $k\epsilon^{II}$ must not be taken much smaller than unity.

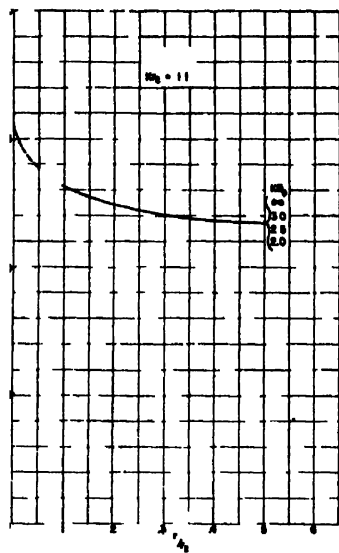


FIG 6a

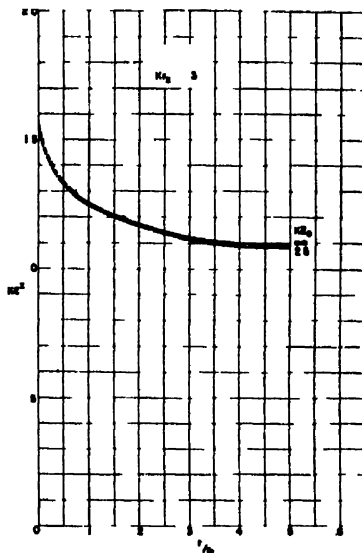


FIG 6b

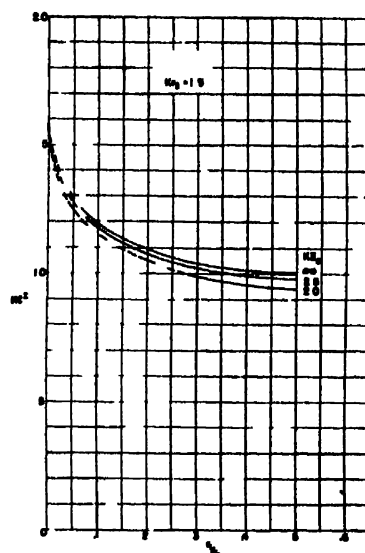


FIG 6c

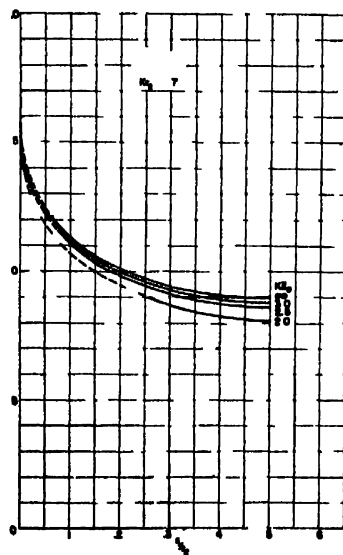


FIG 6d

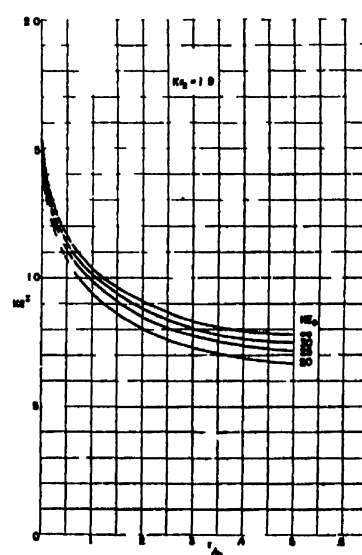


FIG 6e

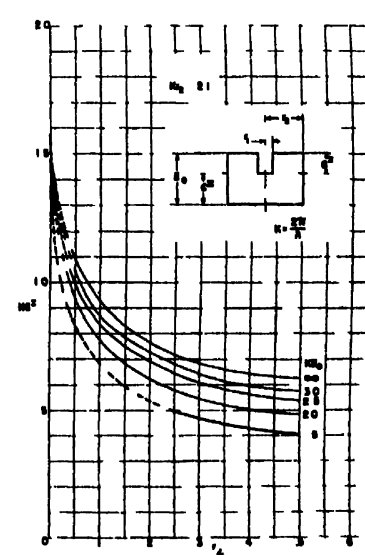


FIG 6f

for other values of r_1/r_2 and obtain by interpolation curves for ke^I vs ke^{II} . On 5a-f the curves for $r_1/r_2 = 2$, indicated by the dotted lines were

plotted in this manner. A check was obtained on the interpolation by a direct calculation for $kr_2 = 1.9$

The Study of a Certain Type of Resonant Cavity and Its Application to a Charged Particle Accelerator

E. S. AKELEY

Department of Physics, Purdue University, Lafayette, Indiana

(Received November 20, 1945)

A linear accelerator for charged particles, consisting of a number of cavity resonators placed end to end with the inside end plates removed, and thus forming a wave guide with two closed ends, is discussed. The shape of each section must be so chosen that the velocity of the particle will correspond to the phase velocity of the radiation of the cavity. In order to determine possible shapes for the cavity, the stationary TM_{01} modes between two parallel conducting planes are determined. In Part 1, excitation of a single mode is discussed, including the shape of the conducting surface that can be used to connect the planes and thus form a cavity. Three cases are considered—when the phase velocity in the cavity is greater than, equal to, and less than that of light. If only one mode is excited,

and the phase velocity is not greater than that of light, then the radius of the cavity becomes infinite at the two plates. In Part 2, the simultaneous excitation of two modes is considered. The phase velocity of one mode is that of light; the other has a phase velocity $1/n$ that of light, where n is an odd integer. The cavity can be made finite by a suitable choice of the relative amplitude of the two modes, and the radius of one end of the cavity. This requires a rather elaborate discussion. This case is of particular interest for the acceleration of electrons. It is shown, furthermore, that only the first mode contributes to the acceleration of the electron. In Section 4, the application of the methods used in Part 3 to the case of an accelerator for heavy particles is touched upon.

INTRODUCTION

IN the electron accelerator patented by W. W. Hansen in 1937, energy is imparted to an electron by letting it pass a number of times through a cavity resonator. J. R. Woodyard¹ thought of modifying this arrangement by placing a number of such resonators end to end and removing their inside end plates. He showed that it should be possible in this way to produce electrons of many hundred million volts electron energy. The shape of the cavity in such an accelerator must be such that the phase velocity of the radiation along the axis corresponds to the velocity of the electrons. Furthermore, the shape should be chosen to give as small as possible a rate of dissipation of energy caused by the finite conductivity of the walls of the cavity. This paper is concerned with the first of these two problems.

PART 1

If one stipulates rotational symmetry relative to an axis, and a finite field on this axis, the stationary TM_{01} modes for the region between two infinite parallel planes of infinite conductivity are of the form

$$\begin{aligned} E_z &= \cos h_n z J_0[(k^2 - h_n^2)^{1/2} r], \\ H_\phi &= \cos h_n z \frac{J_1[(k^2 - h_n^2)^{1/2} r]}{(k^2 - h_n^2)^{1/2}}, \\ E_r &= h_n \sin h_n z \frac{J_1[(k^2 - h_n^2)^{1/2} r]}{(k^2 - h_n^2)^{1/2}}, \end{aligned} \quad (1-1)$$

with $h = n\pi/L$, where n is a positive integer and L the distance between the plates. (z, r, ϕ) are cylindrical coordinates, z being the distance along the axis, r the perpendicular distance from the axis, and ϕ the azimuth. $Ee^{+i\omega t}$ and $He^{+i\omega t}$ are the electric and magnetic field strengths, respectively. h_n is 2π divided by the wave-length in the guide, and k is 2π divided by the wave-length of radiation of the same frequency in space. Each n defines a mode, and these modes can exist independently. However, if the planes are connected by a surface of revolution of infinite conductivity, the relative intensity of the modes excited will be determined by the requirement that the tangential component of E and the normal component of H are zero on the surface, whose differential equation is

$$E_\phi ds + E_r dr = 0. \quad (1-2)$$

This equation states that E is normal to the surface. The requirement that H be tangential to the surface is automatically satisfied as H_ϕ is

¹Private communication, to be published soon.

the only component of II different from zero. First, let us determine the shape of the bounding surface required for the excitation of a single mode. It is convenient to introduce the notation $hz = x$ and $hr = y$ which normalizes the unit of distance relatively to the wave-length. Also let $y_0 = y(\pi/2)$ and $y_0 = y(0)$. There are three cases to consider:

Case 1. $k^2 > h^2$.

$$\begin{aligned} E_z &= \cos x J_0 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y \right], \\ E_r &= \frac{h \sin x}{(k^2 - h^2)^{\frac{1}{2}}} J_1 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y \right], \end{aligned} \quad (1-3)$$

and

$$\begin{aligned} J_0 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y \right] \\ = J_0 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y_0 \right] [\sin x]^{h^2/k^2 - h^2} \end{aligned} \quad (1-4)$$

as the equation of the bounding surface. From (1-4) it follows that

$$J_0 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y_0 \right] = 0.$$

There are two cases to be considered:

$$(a) J_0 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y_0 \right] = 0, \text{ so that } y = y_0(x). \text{ This}$$

is the ordinary cylindrical cavity resonator oscillating in the TM_{01} mode.

$$(b) J_0 \left[\left(\frac{k^2}{h^2} - 1 \right)^{\frac{1}{2}} y_0 \right] \neq 0, \text{ where } y \text{ is no longer}$$

constant but variable as determined by Eq. (1-4).

Cases (a) and (b) represent the same mode of oscillation, only the bounding surface is different.

Case 2. $k^2 < h^2$.

$$\begin{aligned} E_z &= \cos x I_0 \left[\left(1 - \frac{k^2}{h^2} \right)^{\frac{1}{2}} y \right], \\ E_r &= \frac{h \sin x}{(h^2 - k^2)^{\frac{1}{2}}} I_1 \left[\left(1 - \frac{k^2}{h^2} \right)^{\frac{1}{2}} y \right] \end{aligned} \quad (1-5)$$

with the equation

$$\begin{aligned} I_0 \left[\left(1 - \frac{k^2}{h^2} \right)^{\frac{1}{2}} y \right] \\ = I_0 \left[\left(1 - \frac{k^2}{h^2} \right)^{\frac{1}{2}} y_0 \right] [\sin x]^{-h^2/h^2 - k^2} \end{aligned} \quad (1-6)$$

for the bounding surface.

Case 3.

$$k^2 = h^2, \quad E_z = \cos x, \quad E_r = \frac{y}{2} \sin x, \quad (1-7)$$

with the equation

$$y^2 = y_0^2 - 4 \log \sin x \quad (1-8)$$

for the bounding surface.

Case 1 cannot be used for a particle accelerator as the phase velocity is greater than that of light. Case 2 is of interest for particle accelerators in general, as the phase velocity is less than that of light. Case 3 is of particular interest for an electron accelerator after the electron has an energy of more than 10^6 electron volts, when it is moving very nearly with the velocity of light. It is to be noted that only in the first case is the surface finite in size. In all cases of interest for a particle accelerator, the radius of the surface becomes infinite at the two conducting planes. In Part 2, it will be shown for Case 3 above that the addition of another mode, if properly chosen, will give a surface of finite radius.

PART 2

Cavities Suitable of High Velocity Electrons

We shall now consider the simultaneous excitation of two modes, $h_1 = k$ and $h_n = nk$, when n is an odd integer. (If n were even, an unsymmetry would be introduced into the shape of the cavity, and it would not be possible to remove the infinity at both ends.) y_0 finite requires that $E_z = 0$ at the point of contact of the surface and plane. This condition determines the relative intensity of the two modes, and one finds

$$\begin{aligned} E_z &= \cos x - \frac{I_0[(n^2 - 1)^{\frac{1}{2}} y]}{I_0[(n^2 - 1)^{\frac{1}{2}} y_0]} \cos nx, \\ E_r &= \frac{y}{2} \sin x - \frac{n}{(n^2 - 1)^{\frac{1}{2}}} \frac{I_1[(n^2 - 1)^{\frac{1}{2}} y]}{I_0[(n^2 - 1)^{\frac{1}{2}} y_0]} \sin nx, \quad (2-1) \\ dy/dx &= -E_z/E_r. \end{aligned} \quad (2-2)$$

An integral of this differential equation for the bounding surface has not yet been obtained. Computations by numerical integration are in progress. However, certain characteristics of the solution can be deduced. For this purpose, the following results are of importance:

$$(a) \quad \left. \frac{1}{x} \frac{dy}{dx} \right|_{x=0} = \frac{1-n^2}{y_0 - \frac{(3n^2-1)I_1[(n^2-1)^{1/2}y_0]}{(n^2-1)^{1/2}I_0[(n^2-1)^{1/2}y_0]}}. \quad (2-3)$$

Therefore, the surface meets the planes at right angles, and the curve $y(x)$ makes parabolic contact with the line $y=y_0$.

$$(b) \quad \left. \frac{1}{\left(\frac{\pi}{2}-x\right)} \frac{dy}{dx} \right|_{x=\pi/2} = -\frac{n+1}{\frac{y_c}{2} + \frac{n}{(n^2-1)^{1/2}} \frac{I_1[(n^2-1)^{1/2}y_c]}{I_0[(n^2-1)^{1/2}y_0]}}. \quad (2-4)$$

Therefore the curve $y(x)$ makes parabolic contact with the line $y=y_c$ and is convex toward the x axis at $x=\pi/2$.

(c) Let Y_1 be the value of y for which

$$(E_r/x)_{x=0} = 0. \quad (2-5)$$

Then Y_1 satisfies the transcendental equation

$$Y_1 = \frac{2n^2}{(n^2-1)^{1/2}} \frac{I_1[(n^2-1)^{1/2}Y_1]}{I_0[(n^2-1)^{1/2}Y_1]}. \quad (2-6)$$

Also let Y_2 be the value of y for which

$$\left. \frac{1}{x} \frac{dy}{dx} \right|_{x=0} = 0. \quad (2-7)$$

Y_2 satisfies the transcendental equation

$$Y_2 = \frac{3n^2-1}{(n^2-1)^{1/2}} \frac{I_1[(n^2-1)^{1/2}Y_2]}{I_0[(n^2-1)^{1/2}Y_2]}. \quad (2-8)$$

It is quite easy to compute Y_1 and Y_2 for any n by successive approximation, and to show that $Y_2 > Y_1$ and that both Y_1 and Y_2 increase monotonically with n .

(d) The values of (x, y) for which $E_s = 0$ are

given by the equation

$$1 = \frac{I_0[(n^2-1)^{1/2}y] \cos nx}{[I_0(n^2-1)^{1/2}y_0] \cos x}. \quad (2-9)$$

Now since $\cos nx / \cos x$ decreases monotonically from 1 to 0 as x increases from 0 to $\pi/2n$, it follows that y increases monotonically from y_0 to ∞ in this interval, while the interval $\pi/2n < x < \pi/n$, Eq. (2-9) is not satisfied by any real value of y .

Another relation that is easily verified is that $y_0 = Y_1$ is the point where curvature of $y(x)$ is equal to the curvature of (2-9). Furthermore, for $y_0 < Y_1$, $y(x)$ lies below (2-9) in the neighborhood of $x=0$, while for $y_0 > Y_1$, $y(x)$ lies above (2-9) in the same neighborhood.

(e) The values of (x, y) for which $E_r = 0$ are given as solutions of

$$y_1 = \frac{2n}{(n^2-1)^{1/2}} \frac{I_1[(n^2-1)^{1/2}y] \sin nx}{I_0[(n^2-1)^{1/2}y_0] \sin x}. \quad (2-10)$$

One can deduce:

(1) If $x=0$ and $y_0 = Y_1$, then $y_1 = Y_1$.

(2) $\left. \frac{dy}{dx} \right|_{x=0} > 1$. Therefore if $y_0 < Y_1$ then

$y_1(0) < y_0 < Y_1$, and if $y_0 > Y_1$, then $y_1(0) > y_0 > Y_1$.

(3) As $(\sin nx)/(\sin x)$ decreases monotonically from n to 0 in the interval from 0 to π/n , it follows that y_1 increases monotonically from a given value $y_1(0)$ to ∞ in the same interval.

Making use of the relationships contained in paragraphs (a), (b), (c), (d), and (e), it is now possible to discuss the character of the curves for various values of y_0 . In particular, it is possible to show, for certain values of y_0 , that $y(x)$ either becomes infinite or multiple valued at some point. The curves, for which this occurs, are of no interest for a particle accelerator. Therefore, as soon as it is found that a curve has such a characteristic, that curve is not discussed any further. We now discuss separately three cases:

Case $y_0 < Y_1$. Figure 1 shows graphically the relationships of y_0 , Y_1 , and the curves $E_r = 0$ and $E_s = 0$ for this case where (1) $y_1(0) < y_0 < Y_1$, (2) $y(x)$ lies below $E_s = 0$ in the neighborhood of

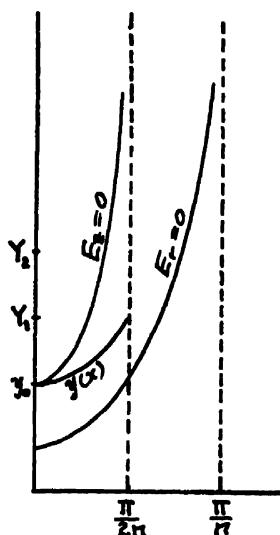


FIG. 1.

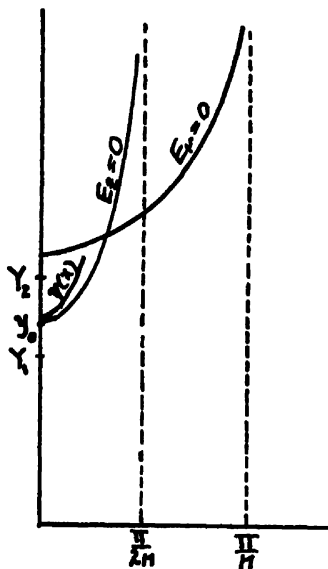


FIG. 2.

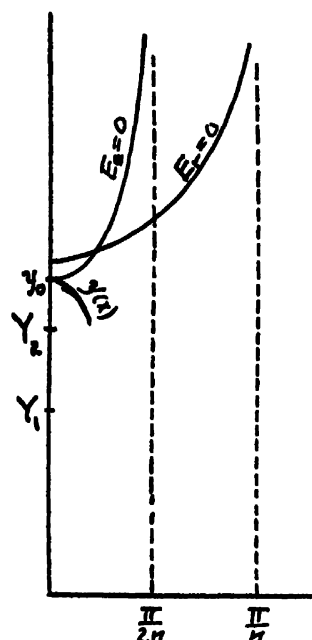


FIG. 3.

$x=0$, and (3) dy/dx is positive in the neighborhood of $x=0$. Now as x increases, dy/dx cannot change sign unless it touches one of the curves $E_s=0$ or $E_r=0$. Now, if it should meet $E_s=0$, it must be horizontal at the point of contact, and that is only possible if $y(x)$ bends backward into a region of negative slope, forming a curve that would not be useful. If it does not meet $E_s=0$, then it must meet $E_r=0$, and if so it must meet the latter tangentially and with infinite slope. This can only occur at $x=\pi/n$. Therefore, $y(\pi/n)=\infty$. Therefore, this case is excluded, not having further interest for us.

Case $Y_1 < y_0 < Y_2$. Figure 2 shows graphically the relationships of the quantities involved. Here $y_1 > y_0 > Y_1$, $y(x)$ lies above $E_s=0$ in the neighborhood of $x=0$, and dy/dx is positive in the same neighborhood. If $y(x)$ should meet $E_r=0$, $dy/dx=\infty$ at contact, and the slope would become negative on crossing this curve. The curve would then bend back toward the y axis. If $y(x)$ does not do this, it must meet $E_r=0$. Then $dy/dx=0$ on crossing, and it will then pass into region of negative slope. There is a possibility that this case would be interesting from the accelerator. If so, it would probably occur for y_0

very slightly greater than Y_1 . A numerical integration will be performed to determine this point.

Case $y_0 > Y_2$. Figure 3 shows the qualitative relationships for this case. The chief difference from the previous case is that dy/dx is negative near $x=0$. It will continue to have a negative slope as x increases until it reaches the line $x=\pi/2$, where it will have zero slope, unless it touches one of the lines $E_s=0$ or $E_r=0$. If it does not meet one of these lines, it would be a suitable shape for a cavity, in the latter case probably not. It would seem to be very difficult to determine necessary and sufficient conditions for obtaining a suitable curve in this case. However, such curves must exist. If y_0 is chosen only slightly greater than Y_2 , $y(x)$ will have a large negative slope. After y has diminished about four units in value, the shape of the curve for greater values of x , would be very close to the type (1-8), and will therefore behave properly. On the other hand, if y_0 is much greater than Y_2 , then the curve will start out flatter, and it may meet parts of the graphs of $E_s=0$ or $E_r=0$, in regions where $x > \pi/n$.

Therefore, for suitable shapes for a cavity resonator for an accelerator, to choose y_0 slightly

greater than Y_2 is most promising, but there may also be possibilities for y_0 slightly greater than Y_1 and for y_0 considerably greater than Y_2 , but these are the only possibilities with only two modes excited.

Let us in conclusion to this section consider the influence of one of these modes on an electron moving along the axis with the velocity of light. The equation of its motion is $z=c(t-t_0)$. The energy imparted to the electron per unit charge in its passage between the two planes is

$$\begin{aligned}\int_0^L dz E_z &= \int_0^L dz e^{-i\omega t} \cos \frac{n\pi z}{L} \\ &= e^{-i\omega t_0} \int_0^L dz e^{-ikz} \cos nkz \\ &= 0 \quad (n \neq 1) \\ &= e^{-i\omega t}/2 \quad (n=1).\end{aligned}$$

It therefore follows that only the mode $h=k$ imparts energy to the electron. The other modes, however, will dissipate energy in the form of heat, caused by the finite conductivity of the walls. One of the considerations in choosing a suitable shape will be to keep this loss as low as possible.

PART 3

Cavities Suitable for Low Velocity Particles

The excitation of the modes h and k with $h > k$, which is of interest when the velocity of the particle is appreciably less than that of light can be treated in the same way as for high speed particles in Part 1. Equations corresponding to (2-1)–(2-4) are easily obtained, and it should not be difficult to determine suitable types of cavity. The shape of each cavity would be determined by the velocity of the entering particle and the amplitude of excitation of that cavity. However, the velocity of the particle and the shape and size of the cavity would change progressively as one passes along the wave guide. Furthermore the velocity of the particle entering any cavity would depend upon the amplitude of excitation of the other cavities, and at the same time the relative excitation of the cavities would depend upon their relative sizes. Thus the accelerator for low velocity particles would appear to have more difficulties both theoretical and experimental connected with it, than are involved in the accelerator for high speed particles.

It might furthermore be interesting to excite more than two modes. Each mode introduces an additional arbitrary constant.

A Millimeter-Wave Reflex Oscillator

J. M. LAFFERTY

General Electric Research Laboratory, Schenectady, New York

(Received July 10, 1946)

The factors which affect the frequency limitation of reflex oscillators are discussed. General design considerations are given for the construction of oscillators of the shortest possible wave-length. The performance of several tubes is given. The highest frequency obtained was 72,000 megacycles.

INTRODUCTION

IN the early part of 1942 Dr. H. V. Neher of the M.I.T. Radiation Laboratory made the first reflex oscillator¹ which operated at centimeter wave-lengths. After that several laboratories placed considerable effort on the design of microwave reflex oscillators in this wave-length region. During the construction of these tubes it was realized that an upper frequency limit had not yet been reached. The material which follows describes the results of experiments performed in an effort to obtain oscillations of the shortest possible wave-length.

In the design of reflex oscillators at extremely high frequencies, the electron gun is of prime importance. The purpose of the gun is to focus electrons from all points of an extended source into a beam of small diameter. The maximum frequency at which the reflex oscillator may be expected to operate is limited by the current density of the electron beam at the position where it passes through the cavity resonator. It can be shown that if the resonator has negligible external loading, the starting current density is proportional to the five-halves power of the frequency. Therefore, the main objective is to design a gun which will transmit as much current as possible through a given diameter aperture for a given beam voltage. Definite upper limits for the value of this current are set by several factors,² the three most important of which are

thermal emission velocity of the electrons, aberrations of the gun lens system, and mutual repulsion (space charge) of the electrons.

From thermal velocity considerations it can be shown that the radius of curvature of the cathode should be made small and the beam angle large to obtain large beam current densities. Just how far one can go in this direction depends primarily on the spherical aberration of the lens system. The use of high beam voltages and cathodes with high emission densities at low temperatures also tend to increase the beam current density. The use of accurately focused wide angle beams would seem to show greatest promise for improvement.

The electric field is greatest in the reflector space. In the absence of appreciable space charge it has a value given by

$$E = 20,240(V_0^{1/2}/\lambda\theta), \quad (1)$$

where E is in volts per centimeter if V_0 is the beam voltage, λ the wave-length in millimeters, and θ the transit time of the electrons in the retarding space expressed in cycles.

Having increased the beam voltage near the point where breakdown occurs, the current density in a given beam is limited primarily by the current density at the cathode surface. Anything which can be done to increase the emission current density of cathodes will extend the upper frequency limit of the reflex oscillator.

To obtain the shortest waves possible from the reflex oscillator, the proportions of the resonator gap must also be given careful consideration. The minimum beam current required to start oscillations is obtained when the beam voltage and gap spacing are adjusted so that the gap transit angle is approximately one-half cycle. For this case, the condition for self-

¹W. C. Hahn and G. F. Metcalf, *Proc. I.R.E.* 27, 106 (1939). H. E. Hollman and A. Thoma, *Hoch. Tech. u. Elek.* 56, 181 (1940). A. E. Harrison, *J. App. Phys.* 15, 709 (1944). J. R. Pierce, *Proc. I.R.E.* 33, 112 (1945). C. C. Wang, *Phys. Rev.* 68, 284 (1945). E. L. Ginzton and A. E. Harrison, *Proc. I.R.E.* 34, 97 (1946). S. Gvosdover and V. Lopukhin, *J. Phys. U.S.S.R.* 10, 275 (1946).

²For a detailed discussion of these factors see J. R. Pierce, "Physical limitations in electron ballistics," *Bell Sys. Tech. J.* 24, 305 (1945).

starting oscillations becomes³

$$I > 8.04 \times 10^6 \left[\frac{(d/\lambda)^2}{R} \right] \frac{1}{p + \frac{1}{4}}, \quad (2)$$

where I , the beam current, is expressed in amperes if d is the gap spacing and λ the wavelength, both expressed in the same units, R is the equivalent shunt resistance of the resonator and load expressed in ohms, p is an integer, and $p + \frac{1}{4}$ is the number of cycles spent by the electrons in the retarding field space. If it is assumed that the electron gun supplies a sufficiently large beam so that the effective current which passes through the resonator may be taken as proportional to the area of the aperture, then I may be replaced by $\pi r^2 J$, where r is the radius of the aperture in the resonator and J is the current density at the aperture. Making this substitution, relation (2) becomes

$$J > 2.56 \times 10^6 \left[\frac{(d/\lambda)^2}{r^2 R} \right] \frac{1}{p + \frac{1}{4}}. \quad (3)$$

For a resonator of given material and negligible external loading, the expression in the brackets depends only on the dimensions of the gap and associated resonator. For a given wave-length the value of d will be determined by the beam voltage used and there remains the problem of selecting a value of r which will make the product Rr^2 a maximum.

In cases where grids are not used across the apertures in the resonator, the optimum value of r obtained by making Rr^2 a maximum may be too large for best operational performance. As the apertures are increased in diameter the d.c. equipotential surfaces between the cavity and reflector are pushed into the cavity by the negative reflector. This has the bad effect of reducing the effective beam voltage in the modulator gap region. In addition the r-f electric field in the gap is reduced because it fringes out through the openings. As a result of these effects, the effective gap coefficient is reduced. The gap coefficient at the edge of the apertures will be the same as that obtained with grids present. However, calculations show that the

gap coefficient at the center of the beam will drop to nearly 0.5 the edge value for an aperture diameter of one-half cycle.

When grids are used, there are several factors to be considered. The grids should intercept a minimum amount of beam current and at the same time give the most effective electrostatic shielding possible. This is accomplished by using grid wire of extremely small diameter. The wire should also be a good r-f conductor and have a low vapor pressure at high temperatures. The smallest wire which is commercially available and suitable for grids is 0.0003-inch diameter tungsten.⁴ The spacing between the wires used in the grid nearest the reflector must be sufficiently small to insure a uniform transit time for all the electrons passing into the retarding space at the same instant of time. The wires used in the grid nearest the gun should be sufficiently close to prevent the r-f field from fringing badly.

Some consideration must also be given to the maximum power which the grids will dissipate. For a grid wire which operates at a high temperature and is very long compared to its diameter, nearly all the power loss at the center is by radiation. The power loss is thus proportional to the surface and hence to the diameter of the wire. The power received by the wire due to current interception is proportional to its projected area and hence to its diameter. Thus, to a first approximation the operating temperature of the grid is independent of wire size.

It appears from expression (3) that the beam current density required for starting oscillations is reduced if the tube is operated under conditions in which p is large. This corresponds to long transit angles for the electrons in the retarding space. If p is made too large, the bunching action in the retarding space is inhibited by space charge effects. Experimental evidence would indicate that there is little advantage in trying to operate with values of p greater than eight or nine. The maximum power output also drops off as p is increased according to the relation⁵

$$P = \frac{0.40}{p + \frac{1}{4}} V_0 I_0, \quad (4)$$

³J. M. Lafferty, "Velocity-modulated reflex oscillator," to appear soon in Proc. I.R.E.

⁴Available from the Cleveland Wire Works, 1131 Chardon Road, Euclid, Ohio.

⁵Lafferty, reference 3, Eq. (48).

where P is in watts if V_0 and I_0 , the beam voltage and current, are expressed in volts and amperes, respectively. The reason for low power output for large values of p is easily understood. For long transit angles the electrons are bunched to the optimum degree by a low r-f modulating voltage. With such a low voltage, little power can be developed in the load.

GENERAL DESIGN CONSIDERATIONS

Perhaps one of the first factors which must be decided upon before proceeding with the design of a reflex oscillator is the beam voltage. Previous experience has indicated that 2000 volts is about the maximum practical beam voltage, taking into account insulation problems, spark over in the tube, and life limitation of an oxide-coated cathode due to positive ion bombardment. The transit angle expressed in cycles for an electron crossing an r-f gap is

$$\theta = 12.85(d/\lambda) V_0^{1/2}, \quad (5)$$

where d = gap spacing in mils (.001"), λ = wavelength in mm, and V_0 = beam potential in volts. Since the gap spacing varies as the square root of the beam voltage, there is little to be gained by going to high voltages. For example, if $\lambda = 5$ mm, a beam voltage of 1600 volts gives a gap spacing of 7.8 mils for a half-cycle gap. Increasing the beam voltage to 2500 volts makes the insulation problems considerably more difficult, but increases the gap spacing only about two mils. In view of these considerations it was decided to use 1600 volts for the beam voltage.

Another question which has to be decided upon is whether or not grids are to be used at the r-f gap. For a 1600-volt beam at five millimeters the gap coefficient at the center of a gridless gap will drop to less than one-half the edge value if the gap diameter exceeds 8 mils. To focus the beam current through such a small aperture would, indeed, place very stringent conditions on the gun design. It is doubtful if a gun could be designed to deliver sufficient starting current (estimated at 30 ma) within a crossover diameter of 8 mils. It was thus decided to use grids and make the aperture larger.

The grids were made from 0.0003-inch diameter tungsten wires spaced 3 mils apart. The

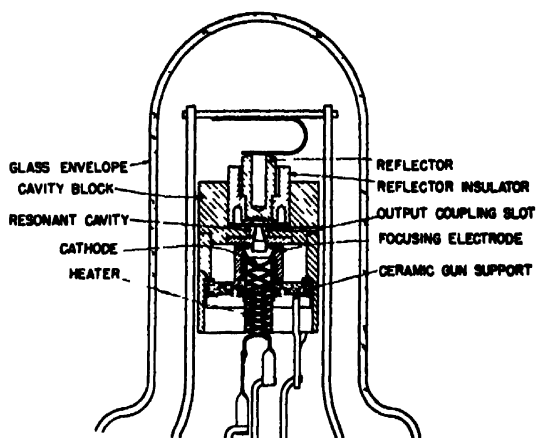


FIG. 1. A millimeter-wave reflex oscillator with a glass envelope. The r-f power radiates from the coupling slot out through the glass envelope.

maximum variation in the transit time of the electrons in the retarding space was found to be less 0.5 percent over the non-uniform part of the field. The effective r-f field strength in the gap is reduced about 25 percent for paths which lie midway between grid wires. The beam power transmission densities required for oscillation would melt the grids if the tube were operated continuously. Pulsed operation therefore was required.

Several types of guns were tried. The ones which performed best are shown in Figs. 1 and 2. The gun shown in Fig. 1 has a special shaped focusing electrode which operates at cathode potential. This gun was designed according to the method described by Samuel.⁶ The gun shown in Fig. 2 has a cylindrical focusing electrode which operates at a negative potential. This potential can be adjusted to give optimum focusing for each tube. Higher beam current densities could be obtained with this gun and it was used in the oscillators which operated below six millimeters. The cathode is $\frac{1}{8}$ inch in diameter and has a radius of curvature of 0.077 inch. It was found that the grids were not damaged or the oscillator performance impaired if the duty cycle did not exceed 0.01. It was found convenient to pulse the beam voltage at a 60-cycle repetition rate with pulses ranging in length from 50 to 160 microseconds. Beam

⁶ A. L. Samuel, Proc. I.R.F., 33, 233 (1945).

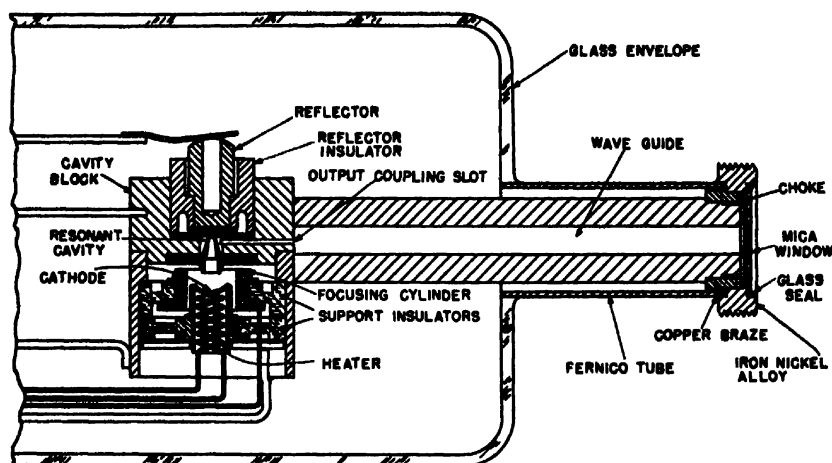


FIG. 2 A millimeter-wave reflex oscillator with a wave guide output. The r f power is transmitted through the non-resonant mica window with negligible loss.

currents with peak values of 50 milliamperes were focused through a 0.028-inch diameter aperture under these conditions with a 1600-volt beam. This represented about 75 percent of the total cathode current.

The reflector consisted of a plane electrode spaced 30 mils from the gap. The aperture near the reflector was made 20 percent larger than the aperture in the cone near the electron gun to accommodate the returning beam. Under an extreme condition such as this where the beam travels less than one diameter into the retarding space, very little is necessary in the way of focusing to insure a large percentage of the beam returning across the gap.

For a 5.5-millimeter oscillator the cavity resonator was 82 mils in diameter and 43 mils high. The coupling slot was 37 mils wide and 20 mils high. The reentrant cone has an included angle of 35° and a 28-mil diameter opening for the electron beam. The gap spacing was 7 mils. For shorter wave-lengths the external dimensions of the cavity, coupling slot, and gap were correspondingly smaller. The dimensions of the cone were kept the same.

TUBE PERFORMANCE

Two types of tubes were constructed as shown in Figs. 1 and 2. In the first type no attempt was made to incorporate a wave guide output or window in the structure. The radiation from these tubes was picked up in a short section of

wave guide placed next to the glass envelope opposite the output coupling slot as shown on the right in Fig. 3. This section of guide contained a wave meter branch, a crystal, and a shorting plunger in the end opposite the tube. An absorption type wave meter was used. Resonance was detected by observing the dip in the crystal current. Wave-length, electronic tuning, starting current, and relative-power output measurements of the various modes could be made. Omitting the wave guide and window on the tube simplified the structure considerably and permitted a large number of experiments to be made on the gun, grids, and cavity with a minimum of effort.



FIG. 3. Set-up for making power and wave-length measurements. A wave guide output tube connected to a bolometer is shown on the left. On the right is shown a tube without a wave guide output in position for making wave-length measurements.

TABLE I. Starting current data.

Tube number	λ	V_0	V_r	I_s	I_0	Relative power
26	9.60	1500	-455	23.9	32.2	100
			-245	20.6		67
			-100	19.8		26
			-35	19.6		7
58	6.05	1500	-810	22.6	56.0	100
			-510	20.7		56
			-310	19.3		26
			-190	18.7		11
			-95	17.4		4
			-20	14.7		2
64	5.45	1500	-370	16.5	20.7	100
			-190	15.2		73
			-60	13.8		50

In order to measure the total power output, a wave guide and window were added to some of the tubes. This structure is shown in Fig. 2 with the gun using the cylindrical focusing electrode.

All tubes were operated by pulsing the beam voltage 60 times a second with 50 to 160 microsecond pulses. The pulsating crystal current thus obtained was fed into a type 208 DuMont oscilloscope. With the crystal used (Sylvania 1N26), a one-inch deflection on the scope corresponded to an r-f power of approximately 0.1 milliwatt.

Table I shows some data taken on the starting current and relative power output for the various modes of oscillation. The more negative reflector voltages correspond to shorter transit times in the retarding space. It will be observed that the starting current and power output both increase with the shorter transit time modes as would be expected from theoretical considerations. The starting current at different wave-lengths cannot be readily compared because of structural variations in the various tubes.

The minimum starting current I_s was determined by operating the cathode at a temperature reduced to the point where oscillations could just be detected. This beam current was then collected and measured as reflector current by operating the reflector a few volts positive with respect to the cavity. The operating current I_0 is the reflector current thus obtained with the cathode operating at normal temperature.

TABLE II. Power output measurements.

Tube number	λ	V_0	V_r	I_s	I_0	P
70	5.80	1750	-40	32	70	12.7
77	5.60	1600	-50	31	48	5.5
78	4.51	1700	-45	38	49	2.6
87	4.15	1700	-350	16	25	0.2

V_0 = beam voltage

V_r = reflector voltage

I_s = peak starting current in ma

I_0 = peak operating current in ma

λ = wave-length in mm

P = peak power output in milliwatts.

The power output obtained from several tubes constructed with wave guide outputs is summarized in Table II. The r-f power was measured with a thermister type bolometer connected across a wave guide section which was coupled to the tube. The wave guide contained a double stub tuner and shorting plunger for impedance matching as shown on the left of Fig. 3. A Type D-162046 unmounted thermister bead was used with an M.I.T. Radiation Laboratory TBN-3EV (Type W) thermister bridge.

The electronic tuning obtained by varying the reflector voltage was found to be about 80 megacycles for most of the tubes below six millimeters.

It was found in all cases that an increase in beam current (by increasing the cathode temperature) resulted in a larger power output. This means that the equivalent shunt impedance of the cavity and load is too low for the available beam current and explains the rather low power output and efficiency obtained at these short wave-lengths. At shorter wave-lengths this condition becomes even more acute and a point is soon reached where oscillations cannot be obtained because of insufficient beam current.

The frequency and power output obtained with these tubes are considerably greater than those obtained by Cleeton and Williams⁷ and Richter⁸ with magnetrons. The reflex oscillator has an additional advantage over the magnetron in that no magnetic field is required. It is doubtful if oscillations of wave-lengths appreciably shorter than four millimeters can be obtained by the velocity modulation reflex principle until new cathodes are developed with considerably higher emission current densities, or some

⁷ E. E. Cleeton and N. H. Williams, Phys. Rev. 50, 1091 (1936).

⁸ H. Richter, Hoch: Tech. u. Elek. akus. 51, 10 (1938).

radically new gun design is invented for producing electron beams of extremely high current densities.

The power output obtained from the tubes described in Table II should be ample for making

dielectric measurements and absorption studies of gases. By making one side of the resonant cavity flexible, it should be possible to construct tubes which may be continuously tuned over a 10 to 15 percent wave-length range.

Resonant Circuit Modulator for Broad Band Acoustic Measurements

GORDON FERRIL HULL, JR.*

Wilder Laboratory, Dartmouth College, Hanover, New Hampshire

(Received July 23, 1946)

A modulation method is described whereby a broad band frequency response is obtained for recording of sound. In particular low frequency sound approaching 0 c p s. can be recorded. The theory of the resonant circuit modulating principle is first discussed followed by a description of the apparatus which was constructed for this purpose.

INTRODUCTION

THE resonant circuit modulator was developed for the purpose of recording sound over a broad frequency band, and particularly to record low frequency sounds. The principle upon which the modulator is based is the following. Consider a series resonant circuit having an inductance L , resistance R , and capacity C connected to a sinusoidal source of constant e.m.f. and of constant frequency. As the capacity C is varied, the current through the resonant circuit will vary and hence the voltage across the capacity will vary. If for the capacity C , we substitute a condenser transmitter and tune the resonant circuit off resonance, then as the capacity of the transmitter is varied by the sound field the voltage across the transmitter will vary according to the shape of the resonance curve. It will be shown that this action produces amplitude (and also phase) modulation of the carrier frequency.

This principle has been suggested by Hans Riegger¹ for converting sound into electrical energy. It is also the general basic principle of the ultra micrometer, a device for the measurement of small displacements, which was devel-

oped by R. Whiddington² and in the Bell Telephone Laboratories by J. R. Haynes.³

The discussion of the resonant circuit modulator and broad band recording system is divided into two parts, the theory of its operation and the experimental apparatus.

THEORY

Series Resonant Circuit, Constant Voltage Source

We begin the analysis of the resonant circuit modulator by considering a series resonant circuit. A constant voltage generator $E = E_0 e^{j\omega t}$ of constant frequency ω and of zero internal impedance is impressed upon the series resonant circuit composed of inductance L , resistance R , and capacity C . We will define C_0 by $1/LC_0 = \omega^2$ as the capacity when the circuit is tuned to resonance, and $Q = \omega L/R = 1/\omega C_0 R$, the circuit Q at resonance. We wish to find the voltage V_c across the condenser as a function of the circuit Q and the capacity of the condenser C . The current flowing in the circuit is

$$i = \frac{E_0}{Z_s} e^{j(\omega t - \phi)}, \quad \text{where} \quad \tan \phi = \frac{\omega L - (1/\omega C)}{R}$$

and

$$Z_s = \left[R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right]^{1/2}$$

* This research was carried out while the author was a member of the Technical Staff of the Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey.

¹ Hans Riegger, *Wiss. Veröff. a.d. Siemens-Konzern* 3, No. 2, 67 (April 16, 1924).

² R. Whiddington, *Phil. Mag.* 40, 634 (1920).

³ J. R. Haynes, *Bell Lab. Record* 13, 337 (1935).

is the impedance of the series resonant circuit. Hence the voltage across the condenser is

$$V_c = \frac{i}{j\omega C} = \frac{E_0}{\omega CZ_s} e^{j(\omega t - \varphi - (\pi/2))}.$$

It is convenient to define $\alpha = C/C_0$ as the ratio of the capacity of the condenser off resonance to the capacity at resonance. Rewriting the equations for Z_s and V_c in terms of α and Q we obtain

$$\frac{Z_s}{R} = \frac{1}{\alpha} [\alpha^2 + Q^2(\alpha - 1)^2]^{\frac{1}{2}},$$

$$\frac{V_c}{E_0} = \frac{Q}{\alpha Z_s/R} e^{j(\omega t - \varphi - (\pi/2))}.$$

Neglecting the time factor

$$\frac{V_c}{E_0} = \frac{Q}{\alpha Z_s/R}. \quad (1)$$

Equation (1), which can be written:

$$\frac{V_c}{E_0} = [1 + Q^2(\alpha - 1)^2]^{-\frac{1}{2}} \text{ when } Q \gg \alpha$$

is plotted independent of Q in Fig. 1 as curve (1). It will be observed that curve (1) has the shape of a resonance curve.

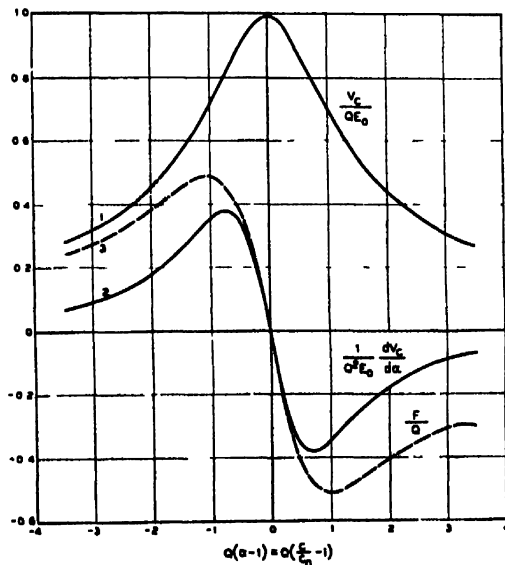


FIG. 1. Generalized voltage-capacity characteristics of resonant circuit.

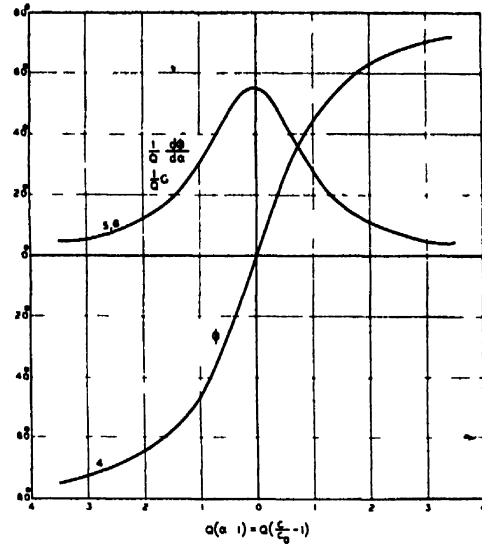


FIG. 2. Generalized phase-capacity characteristics of resonant circuit.

To use the resonant circuit as a modulator, the circuit is detuned from resonance to a point on the resonance curve such that the voltage change produced by a variation of capacity will be a maximum. To find this point it is necessary to find the point at which the slope of curve (1) is a maximum. Consequently we obtain the slope $dV_c/d\alpha$ or what is more convenient to use $(1/QE_0)(dV_c/d\alpha)$. This is

$$\frac{1}{QE_0} \frac{dV_c}{d\alpha} = -\frac{1}{\alpha^2 (Z_s/R)^2} \left[Q^2 \left(\frac{\alpha-1}{\alpha} \right) + 1 \right]. \quad (2)$$

The degree of modulation as a function of α is dV_c/V_c and we obtain for it

$$\frac{dV_c}{V_c} = F \frac{d\alpha}{\alpha} = F \frac{dC}{C}, \quad (3)$$

where

$$F = -\frac{1}{(Z_s/R)^2} \left[Q^2 \left(\frac{\alpha-1}{\alpha} \right) + 1 \right].$$

Equations (2) and (3) when $Q \gg \alpha$ are plotted as curves (2) and (3) in Fig. 1. It is observed from curve (2) that the slope of curve (1) is a maximum at a point on the curve 0.8 of the maximum of curve (1). This can also be shown analytically, by setting $d^2 V_c/d\alpha^2 = 0$, and solving for α in terms of Q . When this value of α is substituted in Eq. (1), $V_c = 0.8QE_0$ is obtained. Hence the

greatest change in voltage for a given change in capacity will occur when the resonant circuit is detuned to 0.8 of the maximum voltage across the condenser. Also in the neighborhood of this point, the slope is changing the least, and, consequently, the greatest linearity of voltage change *versus* change in capacity is to be expected. Although the maximum degree of modulation does not occur at this point, but at a point 0.7 of the maximum of curve (1), the degree of modulation at the 0.8 point is of sufficiently high magnitude.

What has been discussed so far is the amplitude modulation. Phase modulation also occurs since

$$\varphi = \arctan Q \left(\frac{\alpha - 1}{\alpha} \right). \quad (4)$$

It is also desirable to know the slope of (4) and also the value of phase modulation. These are

$$\frac{1}{Q} \frac{d\varphi}{d\alpha} = \frac{1}{\alpha^2 |Z_s/R|^2} \quad (5)$$

and

$$d\varphi = G \frac{d\alpha}{\alpha} = G \frac{dC}{C},$$

where

$$\frac{1}{Q} G = \frac{1}{\alpha |Z_s/R|^2}. \quad (6)$$

Generalized forms of Eqs. (4)–(6) independent of Q when $Q \gg \alpha$ are plotted in Fig. 2. It is not convenient to plot Eqs. (5) and (6) as two separate curves because they differ very little from each other, more specifically by the factor $1/\alpha$.

Before continuing the discussions of the series resonant circuit with constant voltage source, let us consider other ways in which a resonant circuit can be used as a modulator.

Series Resonant Circuit, Constant Current Source

If we substitute a constant current generator $i = i_0 e^{j\omega t}$ for the constant voltage generator in the series resonant circuit, the analysis of its behavior as a modulator is as follows. The voltage across the condenser is

$$V_c = \frac{i}{j\omega C} = \frac{i_0}{\omega C} e^{j(\omega t - (\pi/2))}. \quad (7)$$

Neglecting the time factor,

$$V_c = \frac{i_0}{\omega C}$$

and

$$dV_c = -\frac{i_0}{\omega C} \frac{dC}{C} \quad \text{or} \quad \frac{dV_c}{V_c} = -\frac{dC}{C}.$$

It is evident that amplitude modulation results and no phase modulation. Furthermore the degree of modulation is independent of the resonance curve. In some applications this may be of advantage. However, the degree of modulation is greater in the constant voltage case than in the constant current case by the factor F , and therefore is more desirable to use in spite of the phase modulation which also occurs.

Parallel Resonant Circuit

If a constant current generator $i = i_0 e^{j\omega t}$ of constant frequency ω and of infinite internal impedance is impressed upon a parallel resonant circuit, then the voltage V_c across the condenser is

$$V_c = Z_p i_0 \exp [j(\omega t - \varphi_1 - \varphi_2)],$$

where

$$\begin{aligned} \varphi_1 &= \left[\left(\frac{L}{C} \right)^2 + \left(\frac{R}{\omega C} \right)^2 \right]^{1/2} \\ \varphi_2 &= \left[R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right]^{1/2} \end{aligned}$$

is the impedance of the parallel resonant circuit and

$$\tan \varphi_1 = \frac{R}{\omega L} = \frac{1}{Q}, \quad \tan \varphi_2 = \frac{\omega L - \frac{1}{\omega C}}{R} = Q \left(\frac{\alpha - 1}{\alpha} \right).$$

If $L/C \gg R/\omega C$ and we neglect $R/\omega C$ which is usually permissible, then in terms of Q and α we obtain

$$V_c = \frac{Q^2 R}{Z_s} \frac{i_0}{\alpha} e^{j(\omega t - \varphi)},$$

where

$$\varphi = \varphi_1 + \varphi_2 = \arctan Q \left(\frac{\alpha - 1}{\alpha} \right).$$

Hence we see that a series resonant circuit with

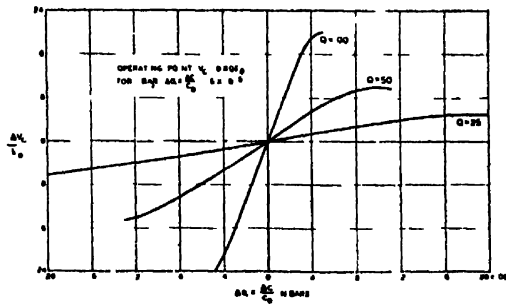


FIG. 3. Voltage-capacity operating characteristics of resonant circuit modulator using type 640 Western Electric condenser transmitter. Change in capacity $\Delta C/C_0$ expressed in equivalent bars pressure

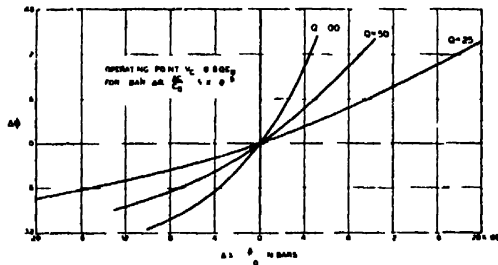


FIG. 4. Phase-capacity operating characteristics of resonant circuit modulator using type 640 Western Electric condenser transmitter. Change in capacity $\Delta C/C_0$ expressed in equivalent bars pressure

a constant voltage source is equivalent to a parallel resonant circuit with a constant current source, provided we substitute $E_0 = QRi_0 e^{j\pi/2}$. Hence all the previous analysis for the series resonant circuit applies to the parallel resonant case.

If, however, a constant voltage source is impressed upon the parallel resonant circuit, then it is evident that the voltage across the condenser will always be a constant irrespective of the manner in which the capacity may change. Hence this circuit will produce no modulation.

Frequency Modulation

If a resonant circuit is used as the frequency determining element of an oscillator, and the capacity of the circuit is varied, frequency modulation (and also amplitude modulation) will result, since for this case $\alpha = 1$, $\omega^2 = 1/LC = 1/LC_0$. Whence

$$2 \frac{d\omega}{\omega} = - \frac{dC}{C} = - \frac{d\alpha}{\alpha};$$

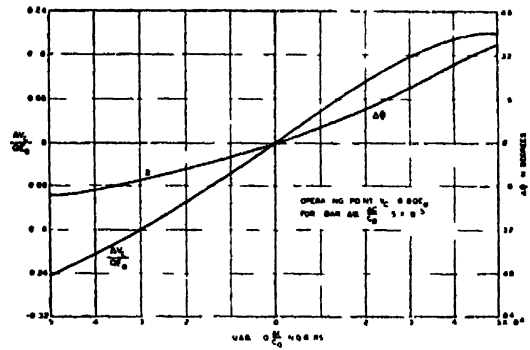


FIG. 5. Generalized voltage-capacity (1) and phase-capacity (2) operating characteristics of resonant circuit modulator using type 640 Western Electric condenser transmitter. Change in capacity $\Delta C/C_0$ expressed in equivalent Q bars pressure.

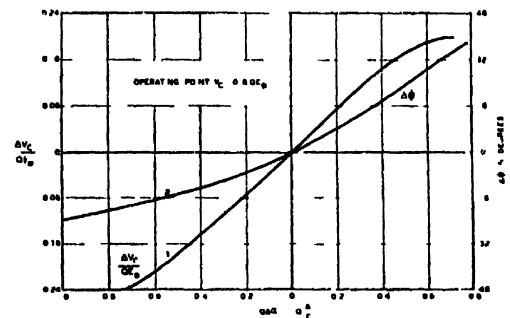


FIG. 6. Generalized voltage-capacity (1) and phase-capacity (2) operating characteristics of resonant circuit modulator for any condenser transmitter.

or twice the fractional frequency variation equals the corresponding capacity variation. Amplitude modulation is also present because V_e is a function of the circuit Q and Q is proportional to ω . The amplitude modulation is not important in this discussion because a limiter circuit would be used to eliminate amplitude modulation.

If the frequency modulated carrier is now impressed on a resonant circuit which is tuned off resonance, then the frequency modulation is converted into amplitude modulation. To show this, define $\omega_0^2 = 1/LC$. Retain the definitions $\omega^2 = 1/LC_0$, $Q = \omega L/R = 1/\omega C_0 R$. Set $\omega/\omega_0 = \beta$ and $d\omega/\omega_0 = d\beta$. Then $(\omega/\omega_0)^2 = C/C_0 = \beta^2 = \alpha$ and $2d\omega/\omega = dC/C = 2d\beta/\beta = d\alpha/\alpha$. Hence frequency modulation is equivalent to amplitude modulation provided we substitute β^2 for α in our previous analysis, Eqs. (1)–(6), and, consequently, the curves in Figs. 1 and 2 can be used for the frequency modulation case.

TABLE I. Operating point $V_e = 0.8QE_0$, $\varphi = \pm 37^\circ$.

Q	Percentage modulation	
	$\Delta V_e/V_e$ for 1 bar	Phase modulation $\Delta\varphi$ for 1 bar
25	.0188%	.014°
50	.0375%	.028°
100	.075%	.056°

Phase Modulation

Evidently the resonant circuit can also be used as a phase modulator. From the curves in Fig. 2 it is seen that the maximum phase shift with capacity change and greatest linearity occur when the resonant circuit is tuned to resonance. The value of phase shift for a given circuit Q and capacity change can be readily calculated from Eq. (6) and the curves in Fig. 2. This phase modulation will also be accompanied by some amplitude modulation as can be seen by referring to curve (1) in Fig. 1.

Application to the Western Electric Type 640 Condenser Transmitter

Modulation

The type 640 condenser transmitter, which is used in the resonant circuit modulator, has a capacity of 50 μf and is normally operated with a polarizing potential of 200 volts. When so used it has a sensitivity of 52 db below 1 volt per bar (dyne per cm^2). A simple calculation shows that the change in capacity, produced by a 1 bar sound field is $\Delta C = 7.5 \times 10^{-4} \mu\text{f}$ and so the fractional capacity change is $\Delta C/C_0 = 1.5 \times 10^{-5}$ per bar. When the transmitter is used in the modulator, and the circuit detuned from resonance such that $V_e = 0.8QE_0$, the curves in Figs. 1 and 2 can be replotted with origin of coordinates at the operating point $V_e = 0.8QE_0$. This has been done in Fig. 3 and Fig. 4 in which $\Delta V_e/E_0$ and $\Delta\varphi$ are plotted against $\Delta C/C_0$ expressed in equivalent bars pressure. The curves are generalized in Fig. 5 for any circuit Q and in Fig. 6 for any circuit Q and for any condenser transmitter.

It is of some interest to know the degree of modulation which will be produced by the resonant circuit modulator using a type 640 condenser transmitter. This can be obtained from Eq. (3) and the curves in Figs. 1 and 3 or by the approximation $F \cong \frac{1}{2}Q$ in $dV_e/V_e = FdC/C$. The

phase modulation can be obtained from Eq. (6) and the curves in Figs. 2 and 4 which give $G = 37Q$ in $d\varphi = GdC/C$. It is of interest to note that the degree of modulation and the phase modulation increase linearly with Q . In Table I are listed the percentage modulation per bar for the type 640 condenser transmitter as well as the phase modulation per bar.

From Table I it is seen that a large percentage modulation is obtained only with large sound fields (of the order of 1000 bars) for circuit Q 's of 25 to 100. For sound fields of the order of 10 bars, Q 's of 100 times as great are required, or else carrier suppression must be used. This will be discussed in a later section.

Distortion

Amplitude distortion produced in the modulator is readily obtainable from the curves in Fig. 3 as well as from curve (1) in Figs. 5 and 6 and is measured by the departure of these curves from a straight line. This gives the total harmonic distortion. Evidently the amount of distortion depends upon the sound field and the circuit Q . The phase distortion is another matter, however, because we are concerned ultimately with the phase distortion in the audio signal. This is given by the ratio of the audiofrequency to the carrier frequency multiplied by the total phase excursion of the carrier, as discussed in a later section.

Response

The audiofrequency response of the resonant circuit modulator is the same as that of the condenser transmitter without d.c. bias. This is so because the amplitude modulation is proportional to the change in capacity of the transmitter, which in turn is a function of the sound field pressure acting on the diaphragm of the transmitter.

Noise

The question of noise produced by the modulator is an important one and can be calculated from the expression⁴

$$V^2 = 4kT \int_{f_1}^{f_2} Z df,$$

⁴ J. B. Johnson, Phys. Rev. 32, 97 (1928); H. Nyquist, Phys. Rev. 32, 110 (1928).

where k is Boltzmann's constant, T the absolute temperature, $Z_r = R/\omega^2 C^2 Z_s$, the real part of the complex impedance Z_s , and $f_2 - f_1$ the band width. To evaluate the noise, some assumptions and approximations will be made. First let us restrict ourselves to a carrier frequency of 4 mc and we will take $C = 50 \mu\text{f}$, the capacity of the type 640 condenser transmitter. Since ω and C are now constants, the inductance L must be a constant and hence R is the only variable in Z_r . If we choose an operating point $V_o = 0.8QE_o$, then $R = 0.8(C/C_o)Z_s$. We can now put Z_r in terms of Q , since $Q = 1/\omega C_o R$, and we have $Z_r = 0.64Q/\omega C$. If Z_r is assumed to be a constant over the frequency band $f_2 - f_1$, which is approximately correct, then $\bar{V}^2 = 2.56kT(Q/\omega C)(f_2 - f_1)$. The frequency band of interest is that in the region of the carrier frequency which ultimately will appear in the rectified audio output, for it is only those noise frequencies in the neighborhood of the carrier frequency which are large in amplitude, because Z_r is large in this frequency band, and which will be transmitted through tuned radiofrequency amplifiers between the modulator and rectifier. Let us assume that the audio band is 10 kc wide, then since the rectifier is a non-linear circuit element, the carrier will beat with those noise frequencies lying in a band ± 10 kc wide on either side of the carrier frequency, thus producing noise in the 10-kc audio band of the rectifier output. Consequently we take $f_2 - f_1 = 20$ kc. The noise voltage $\bar{V}$ can now be calculated. It is desirable, however, to obtain the noise level of the modulator in db referred to a 1 bar sound field. This is $-20 \log_{10} (dV_o/\bar{V})$. Since $dV_o \cong \frac{1}{2}QV_o dC/C$, the noise level becomes $-20 \log_{10} (22.7Q^{\frac{1}{2}}V_o)$ for a carrier frequency of 4 mc a capacity of 50 μf and an audio band width of 10 kc.

In Table II are listed noise levels for three values of Q and one volt across the condenser transmitter ($V_o = 1$ volt). Evidently for sound fields of the order of 1 bar, noise should not be a troublesome factor.

Other sources of noise are also present, but in general they will be small compared to the one we have discussed. Such sources of noise are the load resistance in the rectifier circuit, the input resistance of the d.c. amplifier, and tube noise. These will contribute little to the noise developed

in the modulator, for the latter has been amplified along with the modulated carrier in passing through the radiofrequency amplifier in the receiver. The noise level which has been calculated represents then a minimum noise level for the recording system which, as we shall discuss later, consists of the modulator, a receiver, a d.c. amplifier, and a string oscillograph.

Wave Form of Modulated Carrier

When the resonant circuit modulator is used to produce amplitude modulation, the wave form of the modulated carrier is important, because of the associated phase modulation and, as we will discuss later, because of the desirability of using carrier suppression for low sound fields. Let us assume a sound wave of the form $\cos \rho t$ striking the diaphragm of the condenser transmitter. The capacity of the condenser transmitter will vary according to

$$dC = [dC]_{\max} \cos \rho t,$$

where $[dC]_{\max}$ is the amplitude of the capacity change. The voltage across the condenser transmitter is given by

$$V_o = \frac{E_o}{\omega C Z_s} e^{j(\omega t - \phi - (\pi/2))} = |V_o| \cos(\omega t - \phi),$$

where

$$|V_o| = E_o/\omega C Z_s.$$

Then

$$dV_o = \cos(\omega t - \phi) d|V_o| + |V_o| \sin(\omega t - \phi) d\phi.$$

Now

$$d|V_o| = |V_o| F \frac{dC}{C} \quad \text{and} \quad d\phi = G \frac{dC}{C}$$

from Eqs. (3) and (6). Let

$$F = \frac{[dC]_{\max}}{C} \quad \text{and} \quad \epsilon = G \frac{[dC]_{\max}}{C}$$

Hence

$$dV_o = |V_o| [\delta \cos(\omega t - \phi) + \epsilon \sin(\omega t - \phi)] \cos \rho t.$$

TABLE II. Operating point $V_o = 0.8QE_o = 1$ volt, $\phi = \pm 37^\circ$.

Q	Noise level in db below 1 bar
25	41
50	44
100	47

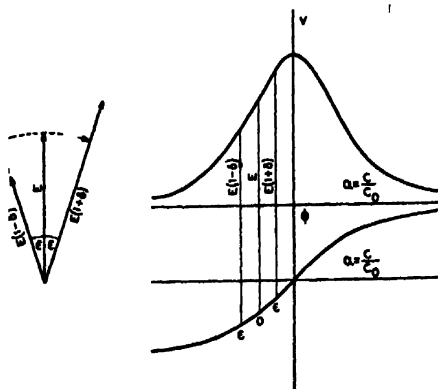


FIG. 7. Vector representation of amplitude and phase modulation of carrier as produced by resonant circuit modulator.

Therefore the wave form of the voltage across the condenser transmitter when a sound wave strikes it is

$$(V_e + dV_e) = |V_e| \{ \cos(\omega t - \varphi) + [\delta \cos(\omega t - \varphi) + \epsilon \sin(\omega t - \varphi)] \cos \rho t \}.$$

This can be written in the form

$$V = E \cos(\omega t - \varphi) + \delta E \cos \rho t \cos(\omega t - \varphi) + \epsilon E \cos \rho t \sin(\omega t - \varphi)$$

or

$$V = E \cos(\omega t - \varphi) + \frac{1}{2} \delta E \{ \cos[(\omega + \rho)t - \varphi] + \cos[(\omega - \rho)t - \varphi] \} + \frac{1}{2} \epsilon E \{ \sin[(\omega + \rho)t - \varphi] + \sin[(\omega - \rho)t - \varphi] \},$$

where we have set $(V_e + dV_e) = V$ and $|V_e| = E$. The first term represents the unmodulated carrier, the second term the amplitude modulation which is in phase with the carrier, and the third term the phase modulation which is 90° out of phase with the carrier. Both modulation terms contain the sum and difference frequencies $(\omega + \rho)$ and $(\omega - \rho)$. This equation holds only for small amplitudes of modulation. For large amplitudes the correct equation is:

$$V = E(1 + \delta \cos \rho t) \cos[\omega t - (\varphi + \epsilon \cos \rho t)]$$

which reduces to the first equation for small amplitudes. The meaning of this equation can be readily understood by means of the vector representation as shown in Fig. 7.

The effect of the phase modulation on the rectified signal of frequency ρ is negligible provided the carrier frequency ω is much higher than ρ and provided the receiver does not cut

off any of the significant side bands of the phase modulation, for the phase modulation of the rectified signal equals the phase modulation of the carrier multiplied by ρ/ω . If the ratio is 0.01, which is reasonable in practice, then a phase modulation of the carrier of $\pm 30^\circ$ will yield only a phase modulation of $\pm 0.3^\circ$ of the rectified signal. In general the phase modulation or phase distortion of the rectified signal can be made as small as desired by increasing the carrier frequency ω .

Carrier Suppression

In general to obtain a reasonable degree of amplitude modulation for low sound fields, it is necessary to suppress some of the carrier after the carrier has been modulated, or to increase the resonant circuit Q to a very high value, which may prove impractical. In view of the phase modulation present amplitude distortion may arise because of suppression of the carrier. In Fig. 8 the vector E representing the carrier is shown modulated in phase through the angle $\pm \epsilon$ and modulated in amplitude from $E(1 - \delta)$ to $E(1 + \delta)$. From these vectors is subtracted a fraction γ of E . The resulting vectors A and B represent the new magnitudes through which the suppressed carrier is amplitude modulated. In order that no amplitude distortion occur, the difference $B - A$ must vary linearly with the sound field pressure for a given value of γ , the degree of carrier suppression. From geometry it is evident that

$$\begin{aligned} A^2/E^2 &= [(1 - \delta)^2 + \gamma^2 - 2\gamma(1 - \delta) \cos \epsilon], \\ B^2/E^2 &= [(1 + \delta)^2 + \gamma^2 - 2\gamma(1 + \delta) \cos \epsilon]. \end{aligned}$$

We are interested in the value of $(B - A)/E$ as a

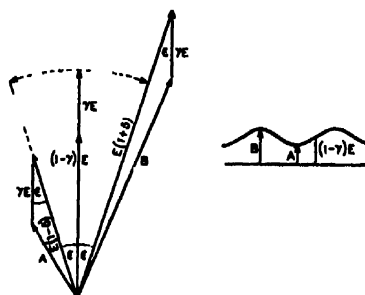


FIG. 8. Vector representation of amplitude and phase modulation of carrier as produced by resonant circuit modulator when carrier suppression is imposed.

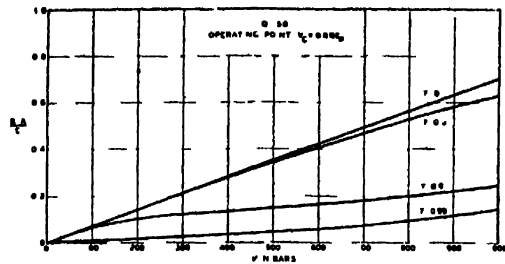


FIG. 9. Magnitude $(B-A)/E$ of amplitude modulation vs. sound field pressure p in bars with various degrees of carrier suppression γ for resonant circuit modulator using type 640 Western Electric condenser transmitter.

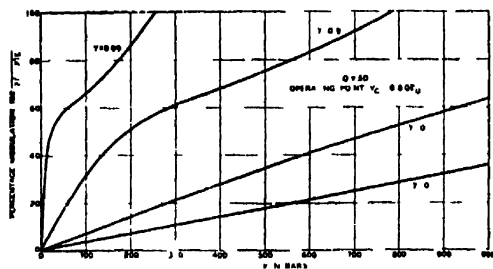


FIG. 10. Percentage modulation vs. sound field pressure p in bars with various degrees of carrier suppression γ for resonant circuit modulator using type 640 Western Electric condenser transmitter.

function of γ and the sound field pressure, and also in the percentage modulation $100(B-A)/[2(1-\gamma)E]$. These have been worked out for the case of the type 640 condenser transmitter in the resonant circuit modulator with $Q=50$ and the operating point $V_o=0.8QE_o$. For this case $F=23.64$, $G=31.51$, $[dC]_{\max}/C=1.5 \times 10^{-5}p$, where p is in bars pressure, and so $\delta=F[dC]_{\max}/C=35.47 \times 10^{-5}p$ and $\epsilon=G[dC]_{\max}/C=47.26 \times 10^{-5}p$. The results are plotted in Figs. 9 and 10. It is seen that for $\gamma=0$ (no carrier suppression) $(B-A)/E$ increases linearly with p and at $p=1000$ bars, the percentage modulation is about 37.5 percent as obtained by the approximate method. For $\gamma=0.9$ (90 percent of the carrier suppressed), the maximum sound field which is permissible without distortion is $p=100$ bars and the percentage modulation for this sound field is about 37.5 percent. And last for $\gamma=.99$ (99 percent of the carrier suppressed), it is seen from the curves that the maximum permissible sound field without appreciable distortion is $p=10$ bars with about the same percentage modulation as before.

BLOCK SCHEMATIC OF BROAD BAND RECORDING SYSTEM

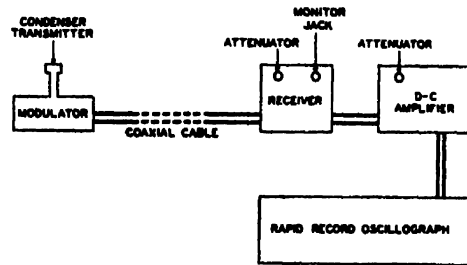


FIG. 11. Block schematic of broad band recording system.

EXPERIMENT

Apparatus

The experimental apparatus used in the broad band recording system consists of the resonant circuit modulator, a receiver, a d.c. amplifier, and a recording string oscillograph. The system is represented by the block schematic Fig. 11

Resonant Circuit Modulator

The modulator is shown in the simplified schematic Fig. 12. The resonant circuit plate load of a conventional crystal controlled oscillator operating at 4 mc is coupled at low impedance (2 ohms) to the modulator circuit. The latter circuit is provided with a coaxial jack to which the condenser transmitter is attached by means of a coaxial line about one foot long inside of a brass tube. The complete resonant modulator circuit then consists of, the inductance L_1 , the variable tuning condenser C_1 of 50 μf , the coaxial line of 25 μf and the type 640 condenser transmitter of 50 μf . The voltage appearing across this total capacity is put on the grid of a tube, and the output is obtained on a coaxial line from a resonant circuit plate load. Measurements on this much of the circuit gave $Q \approx 25$ and 10 percent modulation at 1000 bars sound pressure. This agrees with theory which gives 18.8 percent modulation for $Q=25$ and a sound field of 1000 bars, when it is remembered that the total capacity in the resonant circuit is about 100 μf instead of 50 μf of the type 640 transmitter and consequently the percentage modulation is half as much.

To obtain 10 percent modulation at 100 bars and 10 bars, carrier suppression is necessary.

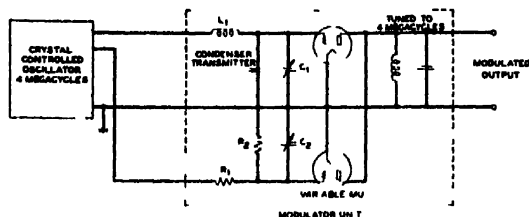


FIG. 12. Simplified schematic of resonant circuit modulator with additional circuit for producing partial carrier suppression.

This was accomplished by means of the lower half of the circuit shown in Fig. 12. Some of the oscillator output which is obtained by suitable coupling is sent through a voltage divider and phase shifter onto the grid of a variable mu tube. The output from the plate of this tube is connected to the plate of the modulator tube, such as to cancel some of the carrier. Cancellation of the carrier is achieved by properly adjusting the phase of the coupling to the resonant circuit plate load of the oscillator and by means of the voltage divider and variable phase shifter which consist of the resistances R_1 and R_2 and the capacity C_2 . The tube gain is controlled by the proper cathode bias. Two bias conditions were used such that for one the carrier was suppressed to 0.1 of its original value and for the other it was suppressed to 0.01 of its value. This gave three operating conditions for the modulator of 10 percent modulation at 10, at 100 and at 1000 bars sound field pressure. For the 1000 bar operating condition the carrier suppression circuit was turned off.

As stated the condenser transmitter was connected to the resonant circuit by means of a coaxial line inside a brass tube. This construction was used in order to prevent the coaxial line from acting itself as a condenser transmitter. On the outside of this brass tube a low pass acoustic filter was constructed which was connected to the back plate of the condenser transmitter in order to allow changes in atmospheric pressure to equalize on both sides of the diaphragm. This filter was about 1 db down at 0.1 c.p.s. If it were not for changes in atmospheric pressure, the air leak in the condenser transmitter could be sealed and then the transmitter

and modulator would respond to frequencies all the way down to 0 c.p.s.

Receiver

The receiver, designed to operate at 4 mc, consists of two r-f stages of amplification having a gain of 30 db per stage followed by a linear diode rectifier. The coupling transformers are broad band (50 kc wide 1 db down) so as not to cut off any of the significant side bands of the phase modulated carrier. A low pass filter connected to the diode rectifier is uniform in frequency response to 8000 c.p.s. and 0.5 ± 0.1 db down at 10,000 c.p.s. A milliammeter was inserted in the diode circuit to measure the rectified d.c. component and also to aid in balancing it out. The maximum distortionless audio output from the receiver when connected to the modulator was 1 volt for sound fields of 10, of 100, and of 1000 bars applied to the condenser transmitter. This voltage is more than sufficient when applied to the input of the d.c. amplifier to drive the string on the rapid record oscillograph to maximum amplitude.

D.C. Amplifier and Rapid Record Oscillograph

The d.c. amplifier was one developed for use with the rapid record oscillograph.⁵ The amplifier is provided with an equalizer such that if the oscillograph string is tuned to 4000 c.p.s., the frequency response of the d.c. amplifier and oscillograph combined is uniform to within ± 0.5 db from 0 to 10,000 c.p.s.

Measurements

Measurements of linearity, frequency response, distortion, and noise were made on the receiver, and on the combined d.c. amplifier and rapid record oscillograph. They were found to be linear and distortionless within the limits of measurement. Their frequency response was uniform to within ± 0.5 db from 0 to 10,000 c.p.s. and their noise level was lower than could be conveniently measured. The measured range of the receiver was 60 db and that of the combined d.c. amplifier, rapid record oscillograph was 30 db. The oscillograph is the limiting element in the system,

⁵ A. M. Curtis, Bell Sys. Tech. J. 12, 76 (1933).

because the limit of resolution of the trace of the string is 30 db.

Measurements of the modulator characteristics were made at the output of the receiver. Known sound fields were applied to the type 640 condenser transmitter by means of a calibrated Western Electric Type 711A Receiver driven by an audio oscillator and coupled to the transmitter. For frequencies below 20 c.p.s. a piston-phone was used. Distortion, linearity, and noise measurements were made using 1000 c.p.s. sound fields of 10, of 100, and of 1000 bars for the three operating conditions. The modulator was found to be linear over a 50-db range for each of the operating conditions within the limits of measurement ± 1 db. The noise level could be determined only in the case of the 10-bar operating condition, and for this case not very accurately. It was found to be greater than 40 db below 1 bar. Distortion measurements for the three operating conditions of the modulator over the 50-db range of each showed that each harmonic component was less than 1 percent of the fundamental. Frequency response measurements are shown in Fig. 13. The response is uniform from 0.5 to 1000 c.p.s. It is down about 1 db at 0.1 c.p.s. because of the low pass acoustic filter associated with the condenser transmitter and has a peak of 5 db at 7500 c.p.s. because of the resonance

of the transmitter diaphragm. This frequency response characteristic is what would be expected for a type 640 condenser transmitter operated without the customary polarizing potential.

These measurements are in good agreement with the theory of the resonant circuit modulator. As the theory shows the modulator should be linear over the straight line portion of the resonance curve and low in distortion. The noise level should be below that of ordinary sound fields encountered and the frequency response should be that of the condenser transmitter.

As a final test the entire system was assembled and records were taken on the rapid record oscillograph of sinusoidal sound fields. These records corroborated the measurements in every way.

CONCLUSIONS

The theory of a resonant circuit modulator employing a condenser transmitter is outlined and it is shown that such a modulator should respond linearly to sound fields and should have low distortion and noise. The frequency response is that of the condenser transmitter used, and with proper construction can be made to cover a wide frequency range and in particular can be made to go to frequencies approaching 0 c.p.s.

Using this modulator, an experimental recording system is described which has a frequency response from 0.1 to 10,000 c.p.s. Measurements made upon this modulator and recording system were found to be in agreement with theory.

I wish to express my indebtedness to Messrs. N. G. Wade, R. T. Jenkins, J. B. Kelly, and Dr. E. C. Wente of the Bell Telephone Laboratories for their valuable suggestions and aid during the course of this work.

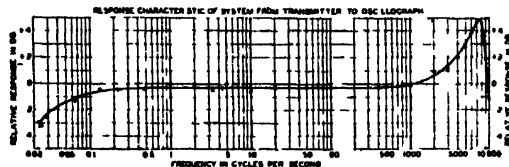


FIG. 13. Frequency response of broad band recording system employing resonant circuit modulator with type 640 Western Electric condenser transmitter.

Transient Temperatures around Heating Pipes Maintained at Constant Temperature

ANDREW GEMANT

The Detroit Edison Company, Detroit, Michigan

(Received June 6, 1946)

This paper deals with the transient temperature distribution caused by conduction around cylindrical heating pipes maintained at definite constant temperatures. The theoretical equation for this case known from the literature is integrated numerically and presented in graphs. The amounts of heat dissipated are also computed and presented. As an application, the influence of pipe diameter upon temperatures obtained at given times is shown for a particular case. For practical installations, the characteristics of which usually deviate from those upon which the theoretical equation is based, the graphs may be used for obtaining close estimates as an aid in designing the installation. It is shown that on two actual installations a satisfactory agreement between measured and computed temperatures (maximum deviation about 10 degrees F) was obtained.

THE PROBLEM

IN connection with the practical problem of melting snow upon sidewalks by means of heated pipes, of which more will be said later, the author has had the opportunity of looking into the theoretical equations describing the process. The basic problem is the characteristics of the transient heat flow caused by conduction into an infinite medium from a cylindrical heat source that is kept at a certain temperature different from the surroundings. This two-dimensional heat flow is described by an equation, discussed in the next section, that contains a definite integral of a certain combination of Bessel functions. It occurred to the author that this integral was not entirely suitable for numerical calculations. While the integral could be expanded in series and integrated by terms, such procedure would have practical use only under certain limited conditions, and not in the whole range of variables encountered. Thus, a numerical integration of the equation and presentation of the results in graphical form seemed desirable.

Such a presentation permits the determination of temperature for any combination of variables. It is strictly valid for one cylinder in an infinite medium. In practical installations these conditions are never exactly fulfilled: There are often several pipes in the ground; also, the extent of the ground in the vertical upward direction is limited. It is thought, however, that even in such cases the theoretical equation and the graphs obtained from it might serve for obtaining a close estimate of expected temperatures and be

of help in the design of the installation. This expectation was verified in the case of two actual installations, as will be shown in a later section of this paper.

The equations and the graphs are valid not only for heating, but for cooling pipes as well.

THE NUMERICAL INTEGRATION OF THE TRANSIENT EQUATION

Let us consider a cylinder of radius a in a medium of thermal diffusivity κ . At a time $t=0$ the pipe is suddenly brought to a temperature corresponding to a temperature difference T_0 above the surroundings and maintained at this level. The question is in what manner does the temperature difference T between any point in the medium and the surroundings depend on the time t and the distance r of that point from the axis of the pipe ($r > a$). The solution was obtained by Carslaw and Jaeger¹ among others and is given by

$$\frac{T}{T_0} = 1 + \frac{2}{\pi} \int_0^\infty \exp(-\kappa u^2 t) \times \frac{J_0(ur)N_0(ua) - J_0(ua)N_0(ur)}{J_0^2(ua) + N_0^2(ua)} \frac{du}{u}, \quad (1)$$

where u is a variable having the dimension of a reciprocal length extending from zero to infinity, J_0 and N_0 = Bessel functions of first and second kind, both of zero order.

¹H. S. Carslaw and J. C. Jaeger, *Phil. Mag.* 26, 473 (1938).

It is advantageous to transform this integral by replacing all quantities occurring in it by dimensionless variables. The substitutions required are the dimensionless temperature ratio

$$\theta = T/T_0, \quad (2)$$

the dimensionless distance ratio

$$\rho = r/a, \quad (3)$$

the dimensionless time function

$$\tau = \kappa t/a^2, \quad (4)$$

and, instead of u ,

$$v = au. \quad (5)$$

With these substitutions, Eq. (1) becomes

$$\theta = 1 + \frac{2}{\pi} \int_0^\infty \exp(-\tau v^2) \times \frac{J_0(\rho v) N_0(v) - J_0(v) N_0(\rho v)}{J_0^2(v) + N_0^2(v)} \frac{dv}{v}. \quad (6)$$

The integration of Eq. (6) was carried out numerically by dividing the total integral in three parts:

$$\int_0^\infty = \int_0^{v_1} + \int_{v_1}^{v_2} + \int_{v_2}^\infty. \quad (7)$$

It can be shown that for v_1 smaller than a certain limit the first term, after expansion in a series, can be solved explicitly, and for v_2 larger than a certain limit the third term can be neglected. It remains then to compute numerically the second term between the limits specified.

Concerning the limit for v_1 , it must fulfill the two conditions: $\tau v_1^2 \ll 1$ and $\rho v_1 \ll 1$. The largest value of τ considered was 10^4 and the largest value of ρ was 40. Hence, $v_1 = 10^{-4}$ is seen to fulfill both conditions satisfactorily.

Using the known expansions² of the Bessel functions, the first integral, multiplied by $2/\pi$, takes the form:

$$-\ln \rho \int_0^{v_1} \frac{dv}{v \ln^2(\gamma v/2)}, \quad (8)$$

where $\ln(\gamma/2) = -0.116$. This can be integrated

directly, giving

$$\ln \rho / (\ln v_1 - 0.116). \quad (9)$$

With regard to the third term, the asymptotic expansion³ of the Bessel functions can be used as an approximation. The integral, multiplied by $2/\pi$, is then represented by

$$-\frac{2}{\pi \sqrt{\rho}} \int_{v_2}^\infty \exp(-\tau v^2) \frac{\sin[(\rho-1)v]}{v} dv. \quad (10)$$

The value of this integral is certainly smaller than the following, obtained by replacing the sin function by unity and v in the denominator by v_2 :

$$-\frac{2}{\pi v_2 \sqrt{\rho}} \int_{v_2}^\infty \exp(-\tau v^2) dv,$$

and the condition required is that this last integral should be small as compared with unity. Introducing the error integral ϕ , the third integral can be neglected if

$$\frac{1}{v_2(\pi \rho \tau)^{1/2}} [1 - \phi(v_2 \sqrt{\tau})] \ll 1. \quad (11)$$

TABLE I. Numerical integration of Eq. (6) for $\rho = 4$ and $\tau = 100$.

v	Fraction below integral involving Bessel functions	Exponential function	$\frac{2}{\pi}$ times integrand	Geometric average of integrands	Average of integrands times Δv
1×10^{-4}	-0.0243	1.00	-155	-118	-0.0118
2×10^{-4}	-0.0282	1.00	-90.1	-77.1	-0.0077
3×10^{-4}	-0.0310	1.00	-65.8	-50.8	-0.0152
6×10^{-4}	-0.0569	1.00	-39.2	-32.4	-0.0130
1×10^{-3}	-0.0420	1.00	-26.8	-20.8	-0.0208
2×10^{-3}	-0.0510	1.00	-16.2	-14.1	-0.0141
3×10^{-3}	-0.0580	1.00	-12.3	-9.77	-0.0293
6×10^{-3}	-0.0730	0.996	-7.75	-6.43	-0.0257
1×10^{-2}	-0.0841	0.990	-5.31	-4.31	-0.0431
2×10^{-2}	-0.114	0.961	-3.49	-3.07	-0.0307
3×10^{-2}	-0.138	0.914	-2.69	-1.98	-0.0594
6×10^{-2}	-0.196	0.698	-1.45	-0.940	-0.0376
0.1	-0.258	0.368	-0.606	-0.117	-0.0117
0.2	-0.386	0.0183	-0.0225	-0.00167	-0.0002
0.3	-0.476	0.000123	-0.000124		
					Sum: -0.3203
					First integral, expression (9): -0.1482
					Total integral: -0.4685
					θ : 0.53

² E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943).

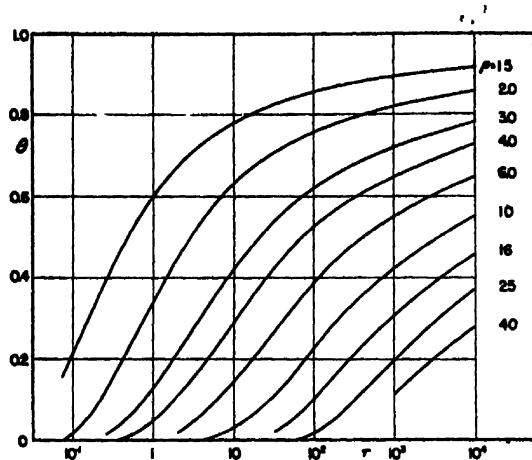


FIG. 1. Temperature as a function of time of heating for various distances from the axis of the pipe.

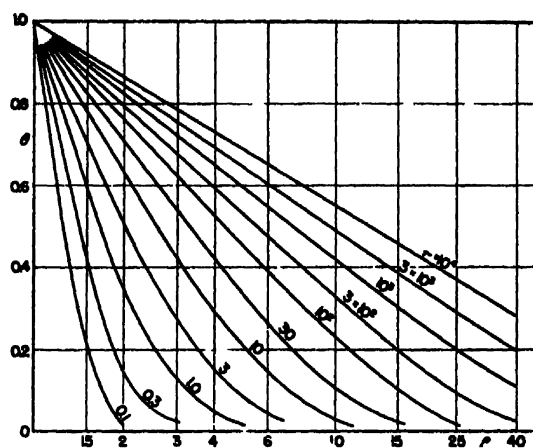


FIG. 2. Temperature as a function of the distance from the axis of the pipe for various times of heating.

From tables² of the error integral it is known that the square bracket in Eq. (11) is very small as compared with unity for arguments equal to or larger than 3. Hence, the limiting value of v_2 is given by

$$v_2 \cong 3/\sqrt{\tau}. \quad (12)$$

The integration of Eq. (6) was thus carried out numerically between the limits $v_1 = 10^{-1}$ and $v_2 = 3/\sqrt{\tau}$; the term (9) was then added to the result of the summation.

Concerning the range of the parameters ρ and τ of the integration, it should be recalled that κ for various materials varies between about 10^{-3} and 10^{-1} . The heating times that might be encountered in practical applications vary between, say, 10 minutes and a day. The radii of heating pipes range roughly between 1 and 10 cm, and the distances from the axis of the pipe between 1 and 100 cm. These ranges are fairly well, although not completely, covered by the range of ρ between 1.5 and 40 and τ between 10^{-1} and 10^4 .

As an example of the numerical integration, Table I for $\rho = 4$ and $\tau = 100$ is given.

THE USE OF THE GRAPHS

The results of the calculations as outlined in the foregoing sections are presented in Figs. 1 and 2. Figure 1 presents θ as a function of τ for various values of ρ , and Fig. 2 presents θ as function of ρ for various values of τ ; the abscissae of both graphs are logarithmic.

There exists extensive literature for the numerical calculation of heat transients *inside* certain enclosures, including cylinders. Patton³ uses a stepwise graphical computation, Emmons⁴ and Dusenberre⁵ apply approximate numerical methods. Graphs, calculated from the rigorous equations, are applied to the interior of enclosures by Newman,⁶ Olson and Jackson,⁷ and Olson and Schultz.⁸ The present writer has not yet found in the literature the application of this principle to the exterior of a cylinder.

The method of using the two graphs, Figs. 1 and 2, is rather obvious, but a few remarks may be made. From the constants of a given installation and the distance and time for which the temperature is required, ρ and τ are computed from Eqs. (3) and (4). Generally the specific values of these two parameters will not be represented by any of the curves on Figs. 1 or 2. From the particular value of τ , the values of θ for a few values of ρ near the one required are read from Fig. 1 and plotted *vs.* $\log \rho$. From a curve drawn through these points, θ for the particular ρ can then be read. Or the same procedure can be carried out starting from the particular ρ and plotting θ *vs.* $\log \tau$, whichever of the two alternatives appears better suited to

³ T. C. Patton, Ind. Eng. Chem. 36, 990 (1944).

⁴ H. W. Emmons, Trans. ASME 65, 607 (1943).

⁵ G. M. Dusenberre, Trans. ASME 67, 703 (1945).

⁶ A. B. Newman, Ind. Eng. Chem. 28, 545 (1936).

⁷ F. C. W. Olson and J. M. Jackson, Ind. Eng. Chem. 34, 337 (1942).

⁸ F. C. W. Olson and O. T. Schultz, Ind. Eng. Chem. 34, 874 (1942).

the case. From θ the temperature is given by Eq. (2).

INFLUENCE OF PIPE DIAMETER ON TEMPERATURE

A practical application of the graphs given in Figs. 1 and 2 is the evaluation of the effect of the pipe diameter upon temperatures attained at certain times and distances from the center of the pipe. It is obvious that with increasing pipe diameter the temperature at a given point increases, but the question is whether this increase is sufficiently large to justify the installation of larger pipes.

An answer to this question can be given only if all the other quantities involved are known. As an illustration of such a case, some data are presented in Fig. 3. A medium of diffusivity $\kappa=0.01$ is taken, and a distance of 10 cm from the pipe axis is assumed. The question is: How does the temperature after certain heating times vary with the pipe radius? From the various data, ρ and τ are computed and θ is read using Figs. 1 and 2 as explained in the previous section. The result for four different heating times is plotted on Fig. 3.

As can be seen, for the range of radii of 1 to 3 cm the increase in temperature difference is nearly proportional to the increase in radius for a 30-minute heating time, but for somewhat longer times, for instance 1 hour and 30 minutes, the effect is less than proportional: The temperatures for 1- and 3-cm radii being in the ratio of 1.0:1.95.

THE AMOUNT OF HEAT DISSIPATED

Further information that can be obtained by using Eq. (1) is the amount of heat Q that is dissipated per unit time per unit length of the cylinder. It is interesting in this connection to compare the present case of heating pipes with the companion case of cables, as discussed by Sticher and the author,⁹ among others. In the case of cables, the amount of heat dissipated is constant while the temperature, in the absence of a sink, would continue to rise indefinitely. In the present case of a pipe of constant temperature the heat dissipated is theoretically infinite at

the first moment and gradually decreases, in the absence of a sink, indefinitely.

This function, Q vs. time is now calculated from Eq. (1). The heat flow at any radius r is

$$Q_r = -2\pi r g (dT/dr),$$

with g = thermal conductivity of the medium. Using θ and ρ as before

$$Q_r/T_0 = -2\pi r g (d\theta/d\rho). \quad (13)$$

Remembering the general relation for any Bessel function

$$Z_0' = -Z_1,$$

one has

$$\frac{d\theta}{d\rho} = -\frac{2}{\pi} \int_0^\infty \exp(-\tau v^2) \times \frac{J_1(\rho v) N_0(v) - J_0(v) N_1(\rho v)}{J_0^2(v) + N_0^2(v)} dv. \quad (14)$$

Since we are interested in the value of Q_r for $r=a$, we have

$$\frac{Q}{T_0 g} = 4 \int_0^\infty \exp(-\tau v^2) \frac{J_1(v) N_0(v) - J_0(v) N_1(v)}{J_0^2(v) + N_0^2(v)} dv. \quad (15)$$

Carrying out the numerical integration of Eq. (15) for various values of τ following the pattern of Eq. (6), the function in question can be obtained.

It is of interest to note that the same function is obtainable from Fig. 2 without any additional

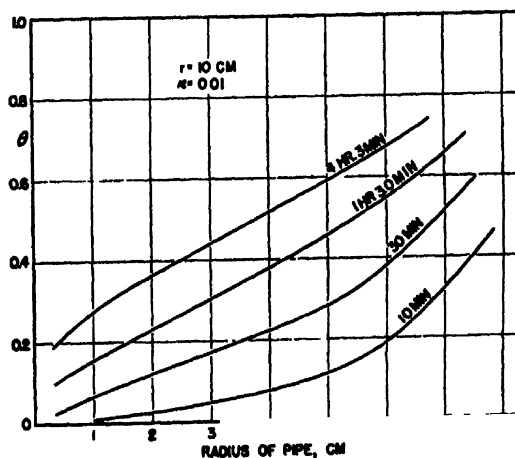


FIG. 3. Effect of pipe diameter upon temperature for various times in a given case, as shown. The notation on the upper curve should read "4 hr. 30 min."

⁹ Andrew Gemant and Joseph Sticher, Trans. AIEE 65, 475 (1946).

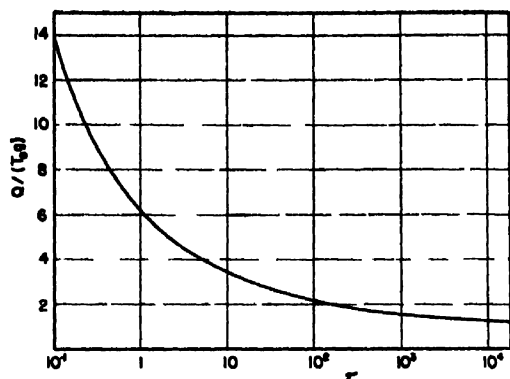


FIG. 4. Heat dissipated per unit time and unit length of pipe as a function of time

computation. From Eq. (13) one has

$$\frac{Q}{T_0g} = -\frac{2\pi}{2.30} \left(-\frac{d\theta}{d \log \rho} \right)_{\rho=1}, \quad (16)$$

and the quantity in the bracket is simply the initial slope of the various curves in Fig. 2.

Figure 4 was constructed by actually calculating numerically the integral in Eq. (15). As a check for the computation, Eq. (16) in conjunction with Fig. 2 was used, the agreement, as was to be expected, was good within 1 percent, the accuracy aimed at for the final results of the numerical computations. In Fig. 4 the dimensionless heat dissipated, $Q/(T_0g)$, is plotted *vs.* τ , from which Q is obtained for any specific case and any given time.

Let us calculate an example. A pipe of 8-cm diameter in soil of diffusivity 0.005, thermal resistivity 100 thermal ohms, and an ambient temperature of 20°C is heated to 100°C. What is the heat dissipated per sec. per cm of pipe length, after 10 minutes and after 1 hour? The values of τ are 0.19 and 1.12, and the ordinates read from Fig. 4 are 10.8 and 6.0, respectively. Thus, with $T_0=80$ and $g=10^{-2}$, the heats dissipated per sec. are 8.6 and 4.8 watts/cm at the two moments in question.

APPLICATION TO A PRACTICAL PROBLEM

A verification on an experimental basis of the graphs given is obviously not necessary since Eq. (1) is based on straight-forward mathematics. It is shown, however, in this section that the graphs can be used for purposes of close

estimates in practical cases in which the conditions underlying Eq. (1) are not fulfilled.

Equation (1) is based on one heating pipe in an infinite medium. In practical cases three deviations from these conditions occur: (a) The pipes are of finite length. This limitation, except for very short pipes, does not seriously affect the applicability of Eq. (1). (b) There are usually several parallel pipes. This circumstance is probably not too serious, either, because the influence of the pipes diminishes rapidly with increasing distance from the pipe. An approximation for not too long heating times may be obtained by simply adding the temperature differences produced by the pipes in the immediate neighborhood of the point in question. (c) The medium in which the pipes are embedded is of limited dimensions. This circumstance will generally invalidate Eq. (1), so that for this reason it cannot be expected to yield other than approximate values

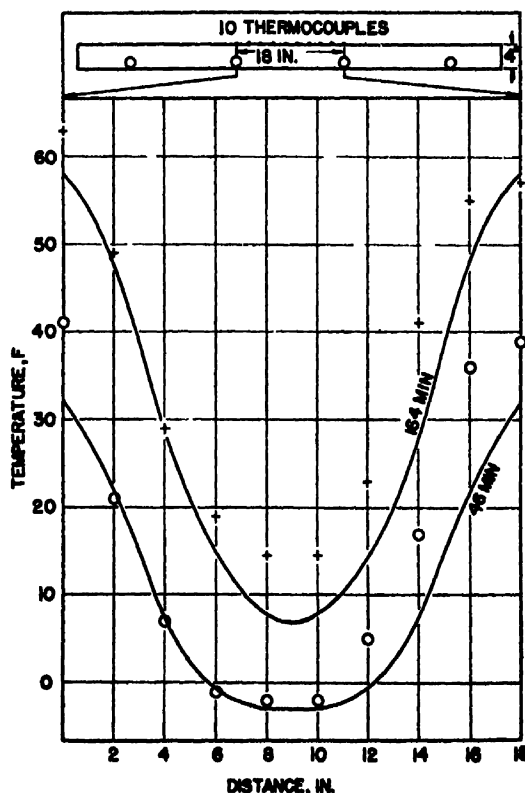


FIG. 5. Temperature distribution on top of a slab containing heating pipes for two different times of heating

○ and + are measured data.
Solid lines are computed from graphs in Figs. 1 and 2

Application of the graphs to the problem of snow melting by means of heat is now shown. The Detroit Edison Company, with which the author is associated, has made some trial snow-melting installations. A cross-sectional schematic diagram of one of these installations is shown at the top of Fig. 5. It is a concrete slab 4 inches deep, simulating a sidewalk, containing four parallel pipes, 1½ in. o.d., spaced 18 inches from center to center. As shown, ten thermocouples were arranged on top of the slab for temperature determinations. The temperature of the surroundings was -3°F and the pipes carried water at 170°F .

In order to apply the graphs, the diffusivity of the embedding medium, concrete, has to be known. The density of concrete was taken as 2.0 g/cc, the specific heat as 0.21 cal./g deg. C, and the thermal conductivity as 0.0029 cal./sec. cm deg. C, as listed for concrete containing 10 percent water; all data were taken from the Landolt-Börnstein Tables.¹⁰ From these data one has $\kappa = 0.007$.

The points given in Fig. 5 indicate the measured temperatures after the water had been flowing for 46 and 164 minutes. The curves were computed from the graphs; the maximum deviation between measured and computed data is about 10°F . From these data the conclusion was drawn that the temperature distribution over the surface was too uneven for the purpose; a similar installation with the pipes on 6-inch centers was then decided upon. Measurements with such an installation, only 12 in. wide, indicated a fairly even temperature distribution; a surface temperature of 32°F , necessary for melting the snow, was reached in about 35 minutes. Computation on such an arrangement also showed even distribution; a temperature of 32°F at a distance from the pipes corresponding to the top surface is reached in about 50 minutes. Considering the relatively small dimensions of the slabs, the agreement between measurements and the estimate from the graphs is satisfactory

¹⁰ Landolt-Börnstein, *Physikalisch-Chemische Tabellen* (Edwards Brothers, Ann Arbor, 1943).

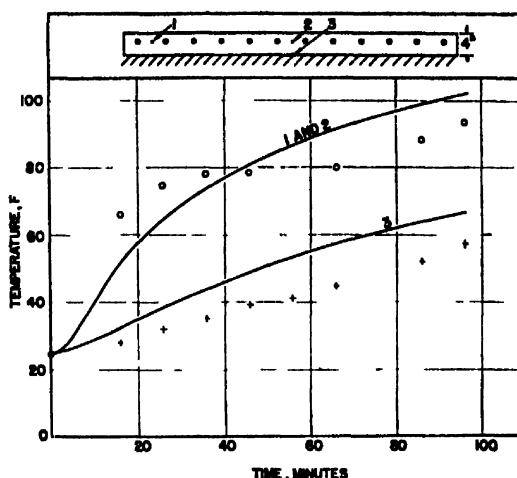


FIG. 6. Temperature as a function of time in a slab containing heating pipes.

Numbers 1 to 3 refer to three thermocouples.
○ and + are measured data.
Solid lines are computed from graphs in Figs 1 and 2.

and shows the usefulness of the latter. Since the installations were in a small cold room, the measured data are higher than the estimates because of the low diffusivity of air.

A cross-sectional sketch of a trial field installation is shown at the top of Fig. 6. The 4-inch deep concrete slab contained twelve iron pipes, 0.84 in. o.d., on 5-in. centers. Three thermocouples were installed as indicated. The temperature of the surroundings was 25°F , that of the water 160°F . The points in Fig. 6 show the measured temperatures as a function of time, the circles being averages of data obtained from thermocouples 1 and 2; the lines represent the estimates from the graphs. The maximum deviation is 12°F . In this case the measured temperatures at the bottom of the slab and partly also near the top are lower than the estimates; this is explainable as a result of the wet ground below the slab and the convection loss caused by wind above the slab. However, even in such an unfavorable case, estimates from the graphs might be useful for designing installations of this kind.

The author is indebted to his colleague, C. M. Weinheimer, for having placed at his disposal the data on the two experimental installations.

The Effect of Temperature on the Strength and Fatigue of Glass Rods

B. VONNEGUT* AND J. L. GLATHART
Preston Laboratories, Butler, Pennsylvania
(Received July 31, 1946)

By the use of an electrodynamic quick-loading device, the strength of scratched soda-lime glass rods was measured at various temperatures between -190°C and 520°C for load durations of 0.1 second, 10 seconds, and 100 seconds. It was found that the strength was highest at the lowest temperature, declined to a minimum at $100^{\circ}\text{--}200^{\circ}\text{C}$, and then increased for higher temperatures. Static fatigue was exhibited at all temperatures, being maximum in the region of minimum strength. Comparisons with the results of other experimenters are made.

INTRODUCTION

THE strength of glass, defined in terms of the stress required to cause fracture, has been a subject of investigation for many years. Most of these investigations have been carried out at room temperature and under ordinary atmospheric conditions. However, glass is frequently called upon to bear large stresses at other temperatures, both lower and higher. As examples we may cite Dewar flasks for holding liquid air, carbonated beverage bottles on ice, and gauge glasses of steam boilers. A knowledge of the way in which the strength of glass depends upon temperature is thus of economic importance as well as being necessary to a full understanding of the fundamental properties and behavior of glass as an example par excellence of a brittle solid.

The experimental results of T. C. Baker, in the years 1938-40, on the mechanical properties of glass have recently been described by Baker, Preston, and Glathart.^{1,2} In 1939, Baker declared that the subject matter seemed to be threatening to escape from the field of physics proper into that of physical chemistry. Accordingly, it became necessary to assign a chemist to the problem, and the senior author's brief connection therewith (1939-40) arose from Baker's feeling. However, the work here described may be said to be straightforward physics. It follows the line of thought reported by Preston³ that, when the

very nature of the problem is obscure, it may be useful to find out how the phenomena vary when the temperature is changed.

The effect of temperature upon the strength of glass has been studied by a number of investigators.⁴⁻¹⁰ The results reported have been somewhat discordant, some tending to show that the strength is independent of temperature,^{4,5} at least up to about 500°C , while other experiments⁶⁻¹⁰ indicate definite temperature effects. This is not surprising, perhaps, when it is remembered that data obtained by different experimenters on the strength of glass at room temperature are not consistent. This is undoubtedly due to the fact that the breaking stress of glass is a function of many variables. Among these are the size and shape of the specimen, the condition of its surface, its chemical composition, and its thermal history, as well as the rate of loading and the water and CO_2 content¹ of the surrounding atmosphere. These are variables which are not always under control in the usual experiment.

The experimental work reported herein was done by the senior author at the Preston Laboratories in 1940, before this country became involved in the war. The changes in personnel and press of war work at the Laboratories prevented publication at the time, and it has only recently been possible for one of us (J.L.G.) to

* Formerly Hartford-Empire Company research fellow at Preston Laboratories. Present address: General Electric Company, Schenectady, New York.

¹ T. C. Baker and F. W. Preston, *J. App. Phys.* 17, 182-188 (1946).

² J. L. Glathart and F. W. Preston, *J. App. Phys.* 17, 189 (1946).

³ F. W. Preston, *J. App. Phys.* 13, 10, 629 (1942).

⁴ H. von Wartenberg, *Zeits. f. tech. Physik.* 13, 479 (1932).

⁵ G. O. Jones and W. E. S. Turner, *J. Soc. Glass Tech.* 26, 35 (1942).

⁶ K. Mengelkoch, *Zeits. f. Physik* 97, 46 (1935).

⁷ M. Eichler, *Zeits. f. Physik* 98, 280 (1935).

⁸ J. S. Gregorius, *Bull. Am. Ceram. Soc.* 15, 271 (1936).

⁹ A. Smekal, *Ergeb. d. exakt. Naturwiss.* 13, 106 (1936).

¹⁰ W. Dawidl and W. Rix, *Zeits. tech. Physik.* 19, 294 (1938).

collect the old data and prepare the present summary.

These results are not intended to be interpreted as the last word on the question of strength *versus* temperature. That must wait for a more adequate appreciation of all the factors involved. However, these data will add just that much more evidence to the view that the temperature coefficient of the strength of glass is by no means negligible.

APPARATUS AND PROCEDURE

The specimens used were annealed soda-lime glass rods six inches long and 7/32 inch in diameter. They were tested in cross bending, between knife edges five inches apart, with central knife-edge loading.

In order to insure a greater homogeneity with respect to surface conditions, the glass rods were uniformly abraded with 00 emery paper applied to them as they were rotated in the chuck of an electric drill.

The apparatus for applying the loads was that used by Baker¹ with the addition of devices for maintaining the specimens at various constant temperatures during tests. Briefly, it consisted of a loud-speaker field magnet, in the magnetic field of which was suspended a light armature coil. This coil was attached to the central knife edge, and when both it and the field magnet were energized by the current from storage batteries, a constant load was applied to the specimen. This load was directly proportional to the armature current, and could be varied from zero to a maximum value of 30 pounds.

The loads were applied in increments of 500 lb./sq. in., and for durations of 0.1, 10, and 100 seconds. The time of the shortest interval was measured with a cathode-ray oscillograph.

The equipment used for making the low temperature tests is shown in Fig. 1. It is essentially a horizontal metal tube which could be surrounded with various low temperature mixtures. This tube contained the test rod, mounted on knife edges. Below the center of the specimen was a narrow vertical tube, extending to the outside of the insulated container, through which the wire stirrup from the armature coil passed.

The two low temperatures used were those of

dry ice and liquid air, respectively, their values being obtained from the literature.

The measurements made above room temperature employed a simple furnace made of steel tubing which could be closed at both ends. These temperatures were maintained constant by a thermoregulator and measured with a thermocouple.

The general procedure in testing a rod at any given temperature was to apply a light load (small armature current) for the chosen time interval. If the rod did not break, the current was increased and a larger load applied for the same time interval. This process was continued until the specimen broke.

Although tests showed that the rods reached temperature equilibrium with the enclosure in six or seven minutes, they were always left in the apparatus for 15 minutes before testing. During this time they were exposed to the atmosphere in the room, which had come to the temperature of the interior of the apparatus. The low temperature equipment was sufficiently airtight so that there was little circulation of the atmosphere through it. No deposit of frost on the specimen was ever observed.

RESULTS

The effect of temperature on the strength of soda-lime glass is shown by the graph of Fig. 2,

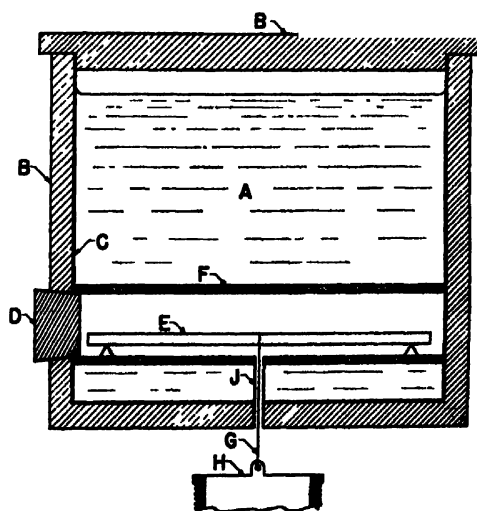


FIG. 1. Low temperature apparatus.

A. Low temperature liquid. B. Celotex thermal insulation. C. Metal container. D. Rubber stopper. E. Glass rod specimen. F. Metal tube. G. Wire loop. H. Armature coil. J. Wire channel.

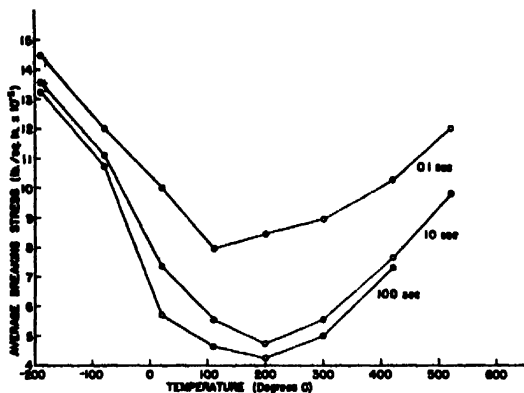


FIG. 2. The breaking strength of glass rods as a function of temperature.

in which the breaking stress in pounds per square inch is plotted against the temperature in degrees centigrade. Each plotted point represents the average for 12 rods, corrected for the loading increment. Since the theoretical form of these curves is unknown, the points are connected by straight lines simply for convenience in associating them.

No determination of the strength of the rods was possible for a load duration of 100 seconds at 520°C. At this temperature, the glass was sufficiently soft that it bent out of shape before its breaking point was reached. For the loads of shorter duration this deformation was not large enough to cause any trouble.

Figure 3 shows the fatigue modulus plotted against the temperature.

DISCUSSION AND COMPARISON OF RESULTS

It can be seen from Fig. 2 that the strength of soda-lime glass, for a 10-sec. duration of stress, decreases from about 13,600 lb./sq. in. at -190°C to about 4800 lb./sq. in. at 200°C. This is a decrease of 65 percent. Beyond 200°C the strength increases, reaching almost 10,000 lb./sq. in. at 520°C. The other two curves follow, in general, the same pattern, with a minimum strength in the 100°-200° range.

Since the chemical composition was the same for all rods used, and since the surface condition was made as nearly uniform as practicable by the abrasion, it is presumed that this temperature effect is due largely to the action of adsorbed moisture. At low temperatures, the weakening

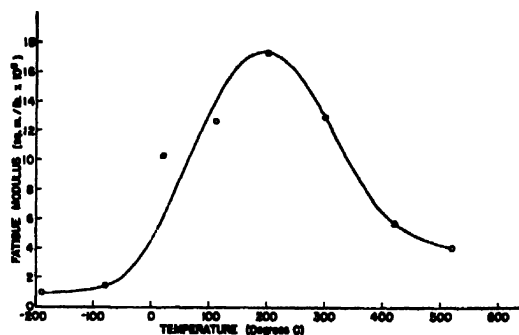


FIG. 3. The fatigue modulus of glass as a function of temperature.

effect due to this moisture layer would be small because of its relative inactivity. Above the 100°-200° range the moisture layer would be driven off. It would be likely to have its greatest effect in the very temperature range in which the glass showed minimum strength. It is undoubtedly true that the severe surface abrasion had an effect on the moisture adsorption. It is known that the heat of adsorption in narrow cracks and crevices of an adsorbent differs greatly from that on the plane surface.

The static fatigue phenomenon, that is, the decrease in strength with duration of load, is shown in Fig. 2 by the positions of the three points corresponding to any one of the temperatures at which measurements were made.

The dependence of fatigue on temperature is shown more clearly, however, in Fig. 3. In this figure the ordinate is the fatigue modulus w_{st_0} defined by Glathart and Preston.³

$$w_{st_0} = \frac{1}{f_{st_0}}$$

Here f_{st_0} is the breaking stress for a duration of load st_0 , e being the base of natural logarithms, 2.718, and t_0 the threshold time, or the least time interval in which a tensile stress, however large, can rupture the material (or the specimen). The fatigue modulus w_{st_0} represents the number of square inches that must be added to the cross-sectional area assigned to carry unit load (one pound) in order to double the life of the specimen.

Figure 3 shows that the fatigue is small at the lowest temperatures, reaches a maximum at about 200°C and thereafter decreases. It is

interesting, and perhaps significant, that the maximum fatigue occurs in the same temperature range in which the strength is least.

Although a smooth curve has been drawn among the plotted points in Fig. 3, there is no reason for believing that it represents the data any better than a number of other curves which might be drawn, including a curve passing exactly through all the points. However, there is some reason for believing that the experimental point at 20°C may be too high, and the line has been drawn accordingly.

It is worth while, perhaps, to compare the fatigue at room temperature of these abraded soda-lime rods with that of the similar unabraded rods used by Baker and Preston.¹ They found that their unabraded rods decreased in strength from 15,900 lb./sq. in. at 0.1-sec. duration of stress to 9100 lb./sq. in. at 100-sec. duration, a decrease of 6800 lb./sq. in. or 42.8 percent. Over the same range, the decrease for the abraded rods of the present report was 4300 lb./sq. in., or 43 percent. This substantiates the conclusion reached by Baker and Preston¹ that scratched glass loses much less strength in a given time interval than does unscratched glass but the percent loss of strength is roughly the same. Their conclusion was based on their results for scratched and unscratched Pyrex glass only. It can now be said to be confirmed even more accurately for soda-lime glass.

The abrading reduced the absolute value of the strength of soda-lime rods at room temperature from 15,900 to 10,000 lb./sq. in. on the 0.1-sec. test, a drop of 37 percent. The corresponding values for the 10-sec. and 100-sec. tests are 29 percent and 38 percent, respectively. Thus abrading alone has the effect of reducing the strength of soda-lime glass about one-third, at room temperature.

The results shown in Fig. 2 are in substantial agreement with those reported by Smekal⁹ who tested scratched Thuringer apparatus glass rods over the same temperature range, using constant rates of loading of 420 and 53 g/mm²/sec. He found that the strength reached a minimum

at 150°C, a decrease of 62 percent from the -190°C value in the case of his lower loading rate.

Mengelkoch⁶ also found a minimum strength at 150°, the decrease from the -190°C value being about 50 percent. Gregorius,⁸ on the other hand, found a minimum at 300°C.

The results reported by Jones and Turner,⁵ however, are in distinct disagreement with those discussed thus far. They used unabraded flat glass laths thoroughly annealed and firepolished. Their experiments covered the range from 20°C to 560°C, and were apparently very carefully controlled. Up to 500°C they found no change in strength with temperature but beyond 500° the strength declined.*

In spite of this, it is believed that the results herein reported give a correct representation of the change in the strength of glass with temperature, under the conditions of the experiment and for the type of specimens used.

CONCLUSIONS

It is concluded that the strength and the static fatigue of scratched soda-lime glass rods exposed to the atmosphere show a pronounced variation with temperature. A minimum of the strength and a maximum of the fatigue at 100°-200° suggest that the form of the curves obtained may be attributed to the effects of adsorbed layers on the surface of the glass. It would be expected that the effects of the layers would be small at low temperature because of the small activity at low temperatures. At high temperatures the effects would again be small because of the evaporation of the layers. Confirmation of this theory will depend on the results of tests on the strength of glass made over the same temperature range in high vacuum, in the absence of adsorbed films.

* Professor Turner states, in a private communication just received, that later measurements by A. J. Holland have been carried down to -40°C, where an increase of modulus of rupture of 20 percent over that at room temperature was found, as well as an increase in impact strength of 17 percent. This is more in agreement with the results reported here.

Plastic Flow, Creep, and Stress Relaxation

Part I. Plastic Flow

CHARLES MACK

Technical and Research Department, Imperial Oil Limited, Sarnia, Ontario

(Received March 28, 1946)

Plastic substances are considered to be composed of units of flow with various yield values. It is shown that in this case the product of the strain rate and viscosity is equal to the sum of the differences between the applied stress and the yield values. This relationship can be applied to any plastic system free of elastic after-effect and expresses their mechanical properties in terms of a coefficient of viscosity which is independent of the stress applied. With the proper choice of the distribution of yield values any kind of relation between stress and strain rate can be established. This relationship is applied to plastic flow which is defined as a deformation mechanism having a curvilinear relationship between stress and rate of deformation and a constant rate of deformation at constant stress. Equations are given for the coefficient of viscosity of such systems and for the relaxation of stress at constant deformation as a function of time.

INTRODUCTION

IDEAL solids and perfect liquids represent two states of matter and two extremes in mechanical behavior. A stress applied to a perfectly elastic body brings about a deformation, the amount of which is proportional to the stress applied and independent of the duration of its application. Because of the symmetric structure of an ideal solid, the deformation involves a change in the distance between the constituent units against chemical forces, and each constituent particle retains identical neighbors.

A perfect liquid can support no other stresses in a state of rest but a hydrostatic pressure. Internal stresses appear as soon as the various volume elements are set in motion relative to each other. These stresses in turn produce a continuously increasing deformation, the rate of which is only determined by the stresses. These stresses are therefore not connected with the magnitude of the deformation at any moment, and they disappear completely, as soon as no further deformation is produced in the liquid. Consequently the deformation of a liquid cannot give rise to any change in potential energy, as in the case of an ideal solid, and the work performed during the deformation is partly used for maintaining the kinetic energy of each moving volume element, and the remainder is converted into heat.

Many industrially important substances have

properties which place them between ideal solids and perfect liquids, i.e., they have rigidity as well as viscosity. The mechanical behavior of such substances is generally referred to as plasticity. The elastic properties of plastic substances are the same as those of ideally elastic bodies except for systems showing the phenomenon of elastic after-effect. Their flow properties, however, deviate from those of perfect liquids in that their relationship between shearing stress and rate of shear is curvilinear. When subjected to a constant stress, plastic substances show two types of plastic deformation as a function of time. In one case the deformation increases linearly with time and the rate of deformation is constant. This type of deformation as a function of time will be termed "plastic flow." The other type has a non-linear relationship between time and deformation, the rate of which decreases with time, and will be referred to as "creep."

The knowledge of the mechanical properties of plastic substances is purely empirical in most cases. It is most desirable to put this knowledge on a more rational basis, since it would make it finally possible to design a plastic substance for a specific use. Much progress has been made by applying the methods of statistical mechanics to molecules.^{1,2} These methods become very com-

¹ H. Eyring and co-workers, *J. Chem. Phys.* 4, 283 (1936), 5, 726 (1937); *J. App. Phys.* 12, 669 (1941); *Ind. Eng. Chem.* 33, 430 (1941).

² A. P. Alexandrov and J. S. Lazurkin, *Acta Phys. Chim. U.S.S.R.* 12, 647 (1941).

plicated when applied to systems containing molecular aggregates of varying size built up from identical molecules or to systems containing such aggregates with several components. Most mechanical tests involve the measurement of a relationship between stress, deformation, and time, and it is a matter of necessity to express such a relationship in terms of exact quantities. For this purpose recourse is taken to the mechanical models suggested by Burgers.³ These models consist of elastic and viscous elements rigidly attached to each other. These elements can be coupled in series or in parallel. No more importance should be attached to these elements than that of mathematical symbols.

The behavior of the simpler forms of these models under stress was discussed by Burgers for the condition of lateral confinement of the models. In the following the derivation of the equations covering the deformation mechanism of plastic substances will be discussed for the condition of tension at constant temperature. The results are directly applicable to the condition of shear, and the shearing stress and shear strain can be calculated from the tensile stress and strain.⁴

GENERAL MODEL FOR PLASTIC DEFORMATION

For viscous flow the Burgers model consists of an elastic spring and a viscous element coupled in series. Spring *A* in Fig. 1 is considered to be perfectly elastic and element *B* consists of a porous piston moving in a cylinder filled with a perfect liquid. A tensile stress, *S*, applied at constant temperature causes an immediate elongation of spring *A* followed by movement of the piston of element *B*. The total strain of the system consists of the sum of the individual strains, consequently the individual strain rates are also additive.

Plastic substances are poly-disperse systems which contain the disperse phase in various degrees of subdivision ranging from individual units of flow to structural units of varying size. A sufficiently low stress produces a strain in the readily deformable part of the system and a resistance

to shear in the structural units. As the stress increases, these units are also deformed at a rate which depends upon their individual yield values. The total strain rate at a given stress is therefore equal to the sum of the strain rates of the individual units.

Such a system can be represented by a model which consists of an elastic element and a number of viscous elements, all coupled in series. Each viscous element has a different yield value. A low stress causes movement of the pistons of all elements with yield values lower than the stress applied. With increasing stress more viscous elements take part in the deformation. Omitting the elastic element, the total strain rate is equal to the sum of the individual strain rates of the viscous elements

$$v = v_1 + v_2 + v_3 + \cdots v_n.$$

The strain rate *v* is $d\epsilon/dt$, where the strain ϵ is preferentially expressed as the "true strain," the natural logarithm of the ratio length after deformation over original length.⁵ With the viscous elements containing the same liquid, the individual strain rates can be replaced by the active stress (applied stress minus yield value) over the

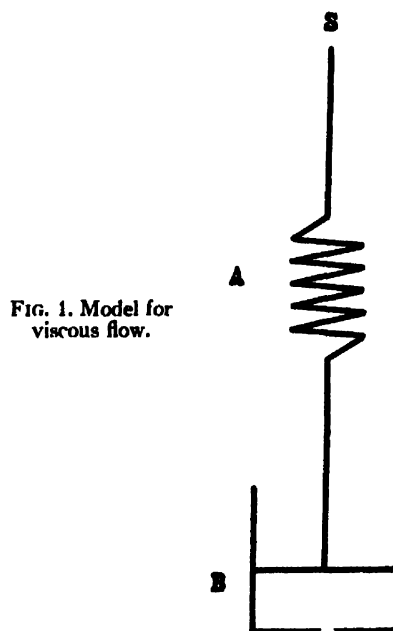


FIG. 1. Model for viscous flow.

³ J. M. Burgers in *First Report on Viscosity and Plasticity* Acad. Sci., Amsterdam 1935.

⁴ A. Nadai, *J. App. Phys.* 8, 205 (1937).

⁵ For details see Markus Reiner, *Ten Lectures on Theoretical Rheology* (Rubin Mass, Jerusalem, 1943).

coefficient of viscosity

$$v = (S - Sf_1)/\eta + (S - Sf_2)/\eta + \cdots (S - Sf_n)/\eta,$$

or

$$v = \sum (S - Sf_i)/\eta. \quad (1)$$

For the condition of tension, the expression η is understood to have the same dimensions as the coefficient of viscosity and to be proportional to it. When the system contains components of different viscosities, the individual strain rates and hence their fluidities ($1/\eta$) are also additive, and the above equation becomes

$$v = \sum [(S - Sf_i)/\eta_i]. \quad (2)$$

These equations can be applied to any plastic system, the plastic deformation of which is not recoverable, and express their mechanical properties in terms of a coefficient of viscosity which is independent of the stress applied. With the appropriate choice of the distribution of yield values any kind of relation between stress and strain rate can be established.

PLASTIC FLOW

Several empirical equations for plastic flow are given in the literature which cover the relationship between stress and strain rate. The most outstanding equations give the stress as a power function of the strain rate and the stress as a function of the logarithm of the strain rate.

(a) Stress a Power Function of Strain Rate

For many disperse systems the curvilinear part of the stress/strain rate diagram conforms to empirical equations which have been given by deWaele,⁶ Farrow,⁷ and Wo. Ostwald.⁸ These equations are similar in form and may be expressed as follows:

$$S/S_0 = (v/v_0)^b, \quad (3)$$

where b is a dimensionless constant, and S_0 and v_0 are constants with the dimensions of stress and strain rate. If the system has a yield value, then S_0 represents the minimum stress which brings about an observable minimum strain rate v_0 .

⁶ A. deWaele, J. Oil Col. Chem. Assoc. 4, 33 (1923).

⁷ F. D. Farrow and G. M. Lowe, J. Text. Inst. 14T, 414 (1923).

⁸ Wo. Ostwald, Kolloid Zeits. 36, 99 (1925).

Written in the following form, $(S/S_0)^{1/b} = v/v_0$, this equation has been criticized by Nadai,⁹ because if the direction of the stress is reversed, a material must deform with the same strain rate in the opposite direction. This condition is only satisfied if $1/b = 1, 3, 5$, etc., but not for fractional irrational or even exponents $1/b$. Further the viscosity $dS/dv = bS_0^{1/b}/v_0 S^{(1-b)/b}$, becomes infinite at $v=0$, $S=0$, which is of course an objectionable conclusion from a physical standpoint. This critique is not justified in view of what has been said.

A system, for which Eq. (3) holds, is deformed at a constant strain rate when a constant stress is applied; hence only the fraction of the system with yield values smaller than the stress applied is deformed. Referring to the model this fraction corresponds to the number of pistons moving. It can be assumed for such a system with one mechanism of viscosity that the summation term in Eq. (1) is proportional to the number of elements deformed and to the applied stress. With n pistons moving

$$\sum (S - Sf_i) = nbS, \quad (4)$$

where b is the proportionality factor which is independent of the stress and has a value between 0 and 1. The distribution of the yield values is here such that their sum is equal to $n(1-b)S$. The term n is equal to v/N where v = number of elements deformed and N = total number of elements per unit volume.

The liquid contained in each element of the model has the same viscosity,

$$\eta = dS/dv_i = (S - Sf_i)/v_i.$$

Hence the differential strain rate is constant for each element, $dv_1 = dv_2 = dv_3$ etc., as can be also demonstrated by differentiating Eq. (1), and the differential of the total strain rate is equal to the differential of the individual strain rate multiplied by the number of elements taking part in the deformation, $dv = ndv_i$, and $dv_i = dv/n$. Combining these terms with Eq. (1) gives

$$\eta v = \sum (S - Sf_i) = v dS/dv/n = nbS \quad (5)$$

and

$$dS/S = b dv/v$$

⁹ A. Nadai and P. G. McVetty, Proc. Am. Soc. Test. Mat. 43, 735 (1943).

which on integration between S and S_0 , v and v_0 leads to Eq. (3), $S/S_0 = (v/v_0)^b$.

Equation (5) contains a constant coefficient of viscosity

$$\eta = nbS/v = n_0bS_0/v_0$$

and since $nS/n_0S_0 = v/v_0 = (S/S_0)^{1/b}$, it follows that

$$\eta = n_0b(S/S_0)^{(1-b)/b}S/v. \quad (6)$$

This equation gives a constant viscosity for plastic flow independent of the stress applied, and contains only S and v as vectors, as $(S/S_0)^{(1-b)/b}$ is equal to n/n_0 and is therefore a scalar. This equation requires the knowledge of the value of n_0 , which can be obtained, if the fractional volume of the elements with strain rate v_0 is known. Otherwise the coefficient of viscosity may be replaced by a constant coefficient of plasticity, η/n_0 .

At a sufficiently high stress the number of moving pistons of the viscous elements can become constant, in which case Eq. (5) may be written

$$\eta v = \sum (S - Sf_i) = nS - \sum Sf_i. \quad (7)$$

The summation term is constant and has the dimensions of stress. Replacing it by f and considering that n becomes 1 leads to

$$\eta = (S - f)/v, \quad (8)$$

which is equivalent to Bingham's equation¹⁰ giving the stress/strain rate relationship for plastic systems with an extrapolated yield value f . It has to be borne in mind that f does not correspond to the highest yield value present in the system but represents the sum of all yield values of the fractions. Equation (8) also applies to the curvilinear relationship between stress and strain rate, in which case, however, f becomes variable.

Equation (3) was derived by considering a system with only one mechanism of viscosity. It can equally well be applied to systems having elements with various viscosities. In the case of two viscosities, for example, we have from Eq. (4)

$$\begin{aligned} v &= v_1 + v_2 = n_1b_1S/\eta_1 + n_2b_2S/\eta_2 \\ v &= v_1 + v_2 = b_1Sdv_1/dS + b_2Sdv_2/dS \end{aligned} \quad (9)$$

from which follows

$$v_1/v_2 = b_1dv_1/b_2dv_2, \quad \text{and} \quad b_1dv_1/v_1 = b_2dv_2/v_2,$$

¹⁰ E. Bingham, *Fluidity and Plasticity* (McGraw-Hill Book Company, Inc., New York, 1923).

hence

$$\begin{aligned} v dS/S &= v_1b_2dv_2/v_2 + v_2b_1dv_1/v_1 \\ &= (v_1^2b_2dv_2 + v_2^2b_1dv_1)/v_1v_2 \\ &= (v_1 + v_2)v_2b_1dv_1/v_1v_2 = (v_1 + v_2)v_1b_2dv_2/v_1v_2. \end{aligned}$$

Adding the last two terms and dividing by $v = (v_1 + v_2)$ gives

$$2dS/S = b_1dv_1/v_1 + b_2dv_2/v_2,$$

and since $b_1dv_1/v_1 = b_2dv_2/v_2$, their sum can be replaced by $2b_1dv_1/v_1$, which on integration leads to

$$S/S_0 = (v_1/v_{01})^{b_1/2} \cdot (v_2/v_{02})^{b_2/2} = (v/v_0)^b. \quad (10)$$

If the system has two values of S_0 with S_{02} larger than S_{01} , then at stresses lower than S_{02} , the relationship between $\log S$ and $\log v$ has a slope different from that at stresses larger than S_{02} . With the number of elements of viscosity η_1 being small in comparison with the number of elements of viscosity η_2 , a measurable deformation may be obtained only at a stress equal to S_{02} . In cases where the system contains several mechanisms of with different values of S_0 , the relationship between $\log S$ and $\log v$ may be curvilinear for stresses smaller than the value of the largest S_0 .

Equations (3) and (10) are based on the assumption that the extensibility of the viscous elements is unlimited. In the case of limited extensibility, a low stress deforms a certain number of elements at a low strain rate. With increasing stress the strain rate of the individual elements increases, but since some of the elements have reached their maximum extensibility, the number of elements taking part in the deformation has decreased. This decrease does not occur suddenly with each increment in stress, but is a gradual process and will be treated in detail in connection with the discussion of creep. For a decrease in the number of elements deformed, which hypothetically depends on the stress only, the ratio of the number of elements deformed, n/n_0 , is smaller than 1. In these cases the active stress is bS as before, however, because of the limited extensibility of some elements, the strain rate is only a fraction of what it should be and is equal to av , where a is a numerical constant such that $0 < a < 1$. Hence Eq. (5) assumes the form

$$\begin{aligned} dS/dv/n &= nbS/av \\ \text{and} \quad (S/S_0)^a &= (v/v_0)^b \end{aligned} \quad (11)$$

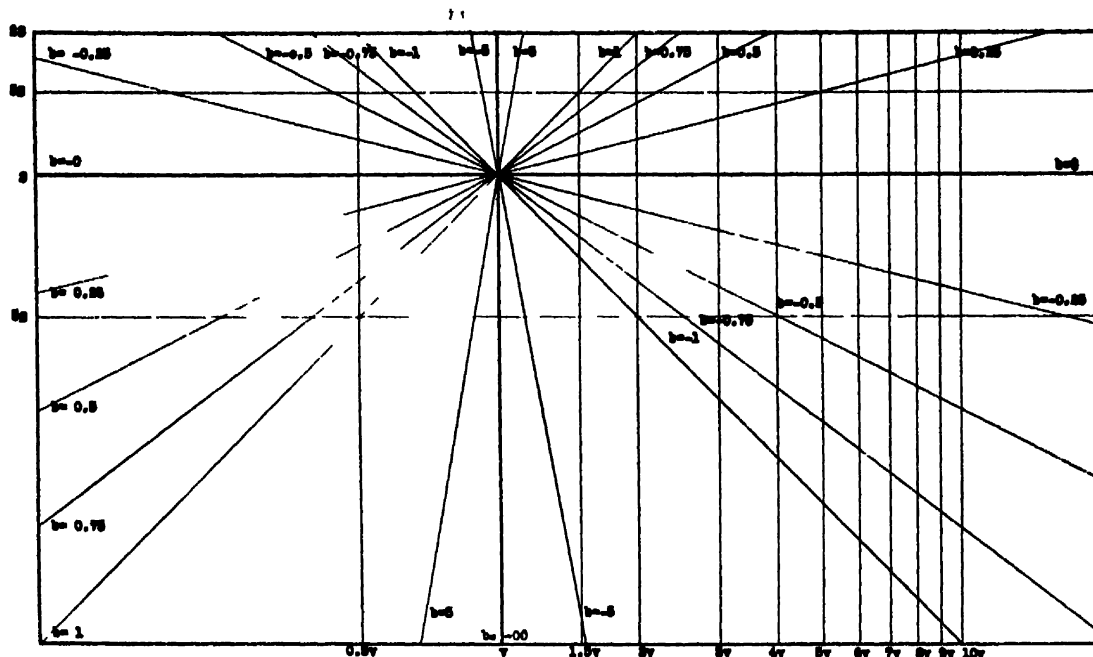


FIG. 2. Relationship between stress and strain rate for various values of b . The notation $b=00$ in the lower center should read $b=\infty$.

or

$$(S/S_0)^{-a} = (v/v_0)^b \quad (12)$$

the latter term holding for systems where dv is negative, i.e., where the strain rate decreases with increasing stress. The viscosities derived from Eqs. (11) and (12) are

$$\eta = \eta_0 (b/a) (S/S_0)^{(1-b)/b} (S_0/S)^{(1-a)/b} S/v, \quad (13)$$

and

$$\eta = \eta_0 (b/a) (S/S_0)^{(1-b)/b} (S_0/S)^{(1+a)/b} S/v. \quad (14)$$

In both cases the ratio η/η_0 is the product of two terms in S , the first expressing the increase in the number of elements deformed with increasing stress and the other representing the increase in the number of elements having reached the limit of their extensibility with increasing stress.

For $b/a > 1$ the total strain rate increases with increasing stress, and for $b/a < 0$ the total strain rate decreases while the individual strain rates increase with increasing stress. In either case an increase in stress causes an increase in the yield value, since, with n elements having reached their limit of their extensibility, the original yield value is increased to that of the $n+1$ st element. Hence

a deformation increases the strength of the system, i.e., the stress that can be supported, and this phenomenon may be termed "stress-hardening."

Systems with negative values of b/a are usually connected with a type of flow termed dilatancy. Actually the mechanical behavior of such systems does not allow for any increase in volume. At a sufficiently high load stress the deformation of all the viscous elements will come to a stop. A further increase in stress will distort the system and cause dilation. An example of this kind of behavior will be given in a separate part.

The relationship between stress and strain-rate is represented in Fig. 2 for values of b and b/a varying between $-\infty$ and $+\infty$. With $b=0$ the stress becomes independent of the strain rate. A substance of such properties does not exist, since at $b=0$, the active stress and the viscosity are zero. As b/a increases from 1 to ∞ the strain rate becomes more independent of the stress. The same also holds for b/a decreasing from 0 to $-\infty$. Keeping η_0 , S , and v constant, the viscosity decreases with increasing values of b and reaches a minimum at $b=1$ and increases with increasing values of b/a and becomes infinite at $b/a = \infty$.

The same also holds for negative values of b/a . With $b/a = +\infty$ or $-\infty$ the strain rate is constant and independent of the stress applied, and the viscosity is infinite. These conditions are met with an ideal elastic body.

(b) Relaxation of Stress

If the system is first deformed and then the strain kept constant, spring A in Fig. 1 contracts and causes a continued motion of the pistons of the viscous elements, until the potential energy stored in the spring is spent. Under this condition the total strain rate is zero and is composed of the strain rate of the elastic spring and the sum of the individual strain rates of the viscous elements, hence

$$dS/Edt + v = 0, \quad (15)$$

where E is the modulus of elasticity of spring A . Substituting v from Eq. (6) and integrating between S and S_0 , the stress at the instant when the strain is kept constant, t and 0 gives

$$(S_0/S)^{(1-b)/b} - 1 = n_0(1-b)(S_0/S)^{(1-b)/b} Et/\eta. \quad (16)$$

If the system has a yield value, the stress will relax during an observable time to a value equivalent to that of the yield value and to zero in infinite time. This equation can also be applied to systems which have different mechanisms of elasticity and viscosity, in which case E/η must be replaced by $\Sigma^1/\eta/\Sigma^1/E$.

The term E/η is constant with the dimensions of $1/t$ and may be replaced by $1/\tau$ where τ is the relaxation time.

Since $n_0 (S_0/S)^{(1-b)/b} = n_0$, the number of viscous elements deformed at stress S_0 , Eq. (16) may be simplified into

$$(S_0/S)^{(1-b)/b} = n_0(1-b)t/\tau + 1. \quad (17)$$

Although the term $n_0(1-b)/\tau$ contains the constant relaxation time, τ , its value depends upon n_0 and therefore upon S_0 . Replacing this term by $1/tr$, where tr is a variable relaxation time depending upon the initial stress, gives

$$(S_0/S)^{(1-b)/b} = t/tr + 1. \quad (18)$$

The variable relaxation time is the time at which the original stress is reduced to a value of $S_0/2^{b/(1-b)}$.

This equation holds only for systems for which

$0 < b < 1$. With $b/a > 1$ and $b/a < 0$, there is no relaxation, since immediately after the strain is kept constant, all viscous elements with yield values equal to or smaller than S_0 have reached their maximum extensibility. Such systems, therefore, behave at constant strain and temperature like ideal solids.

(c) Strain Rate an Exponential Function of Stress

Another equation which covers curvilinear relationships between stress and strain rate has been suggested by Ludwik¹¹ and shows the stress to be proportional to the logarithm of strain rate in accordance with

$$\exp (S - S_0)/S_0 = (v/v_0)^b \quad (19)$$

or

$$\exp (S/S_0) = e(v/v_0)^b, \quad (20)$$

where b , S_0 , and v_0 have the same meaning as before. This equation was mathematically derived by Prandtl¹² by considering a model consisting of a row of molecules which are bound to their normal positions of rest by forces which increase proportionally with the deviation of the molecule from its particular position of rest, i.e., these forces are of an elastic nature.

Applying statistical mechanics to his reaction rate theory, Eyring¹³ arrived at a relationship for plastic flow which is similar to Eq. (19).

$$v = \frac{\lambda kT}{h} \exp (-F^\ddagger/kT) \times [\exp (S\lambda/2kT) - \exp (-S\lambda/2kT)], \quad (21)$$

where λ = the average distance jumped by a molecule, therefore $S\lambda$ = work done by a potential, $-S\lambda$ is the amount of work for a similar motion in the opposite direction, k = Boltzmann's constant, h = Planck's constant, T = absolute temperature, and $F^\ddagger$ = free energy change of going from normal to activated states. If the free energy of activation is large, S must be large before v will have a noticeable value, therefore the second term in the parenthesis is negligible. Introducing

¹¹ P. Ludwik, *Elements der Technologischen Mechanik* (Verlagsbuchhandlung, Julius Springer, Berlin, 1909).

¹² E. Prandtl, *Zeits. f. angew. Math. u. Mech.* 8, 85 (1928).

¹³ H. Eyring, *J. App. Phys.* 4, 283 (1936).

a similar term for v_0 gives

$$\exp (S-S_0)=(v/v_0)^{2kT/\lambda}. \quad (22)$$

The Eyring equation has been obtained for a system containing molecules of large size for which the free energy of activation is high. Yet substances of high molecular weight are known which behave like perfect liquids under stress. A viscous asphalt,¹⁴ for example, having viscosities of 10^7 and 7.07×10^6 poises at 15 and 25°C, has a value of 22700 for $F\ddagger/k$. Equation (22) has also been found to hold for the curvilinear part of the stress/strain rate diagram of disperse systems such as greases, paints, and clay ships,¹⁶ the total free energy of activation of which is relatively small, although it may be high for individual units of the disperse phase. These considerations seem to indicate that the mechanism of plastic deformation for a substance obeying Eq. (19) is similar to that obtained for Eq. (3). Such a system can also be assumed to consist of an elastic element and a number of viscous elements coupled in series for which Eq. (1) holds

$$v=\sum (S-S_f)/\eta.$$

If the summation term is independent of the stress and proportional to the number of the viscous elements deformed, then

$$\eta v=nvdS/dv=n \text{ const.}, \quad (23)$$

where const. has the dimensions of a constant stress. Since S_0 is the lowest stress which produces an observable plastic deformation, bS_0 can substitute for const., and

$$dS=bS_0dv/v$$

which on integration between S and S_0 , v and v_0 leads to Eq. (19). From the above equation the following terms for viscosity are obtained

$$\begin{aligned} \eta &= n_0 b S_0 / v_0 = n b S_0 / v \\ &= n_0 b \exp [(S-S_0)/bS_0] S_0 / v. \end{aligned} \quad (24)$$

The distribution of the yield values is in this case such that their sum is equal to $n(S-bS_0)$.

Combining Eq. (24) with the general Eq. (15) for relaxation at constant strain leads to

$$\begin{aligned} \exp [(S_i-S)/bS_0] \\ = n_0(t/\tau) \exp [(S_i-S_0)/bS_0] + 1 \end{aligned} \quad (25)$$

and with $n_0 \exp [(S_i-S_0)/bS_0] = n_i$

$$\exp [(S_i-S)/bS_0] = n_i t/\tau + 1 = t/t_r + 1. \quad (26)$$

The term, $\tau/n_i = t_r$, represents the time at which the initial stress is reduced to a value of $S_i - bS_0 \ln 2$, and depends on the initial stress in the same way as in Eq. (17). For systems containing several mechanisms of elasticity and viscosity, $1/\tau$ becomes $\sum 1/\eta / \sum 1/E$.

Equations (3) and (19) giving the strain rate as a power function and an exponential function of the stress are applicable to plastic systems of the sol-gel type which are subject to plastic flow and form the basis for more complicated systems. As the latter systems are more frequently encountered, the presentation of data will be reserved for a separate paper. It may be mentioned here that experience so far seems to give preference to Eq. (3) for systems where the structural elements of the substance under test remain in a high degree of disorder. In systems, where the structural units orient themselves to a higher state of order under stress, Eq. (19) appears to be applicable.

¹⁴ C. Mäck, unpublished work.

¹⁶ R. E. Powell and H. Eyring, *Nature* **154**, 427 (1944).

Plastic Flow, Creep, and Stress Relaxation

Part II. Creep

CHARLES MACK

Technical and Research Department, Imperial Oil Limited, Sarnia, Ontario

(Received March 28, 1946)

The general equation covering the deformation of plastic substances given in Part I is applied to creep. Creep is defined as a mechanism of deformation for systems which have a curvilinear relationship between stress and strain rate and a curvilinear relationship between strain and time at constant stress. Creep is connected with changes in the internal structure of a plastic substance and results in an increase in strength of such materials through work-hardening. Equations are derived which give the stress as a function of strain rate and time (time-hardening), as a function of strain rate and strain (strain-hardening) and as a function of strain rate, strain, and time. The difference between time-hardening and strain-hardening is discussed. Expressions are given for the coefficients of viscosity of such systems which are independent of the stress applied. The relaxation of stress at constant strain is discussed, and it is shown that the stress relaxation depends upon the history of the substance under test. The concept of creep is also applied to thixotropic systems which are considered as cases of work-softening.

INTRODUCTION

IN Part I¹ plastic systems of the sol-gel type have been discussed. Gels differ from this type only in degree, in that they contain a larger number of elements of three-dimensional structure. The interlocking of these structural elements gives the system an appearance of rigidity. The formation of three-dimensional structures is caused by attractive forces between the individual molecules which are either of the van der Waals type, or in the case of polar molecules are caused by hydrogen bonds which are caused by the interaction between hydrogen and two other atoms. This bond is highly directional, and therefore resembles to some extent a covalent bond, although the free energy of a hydrogen bond is considerably lower.^{2,3}

Amorphous solids such as thermoplastic and thermosetting plastics are generally obtained by polymerization. They also have a gel-like structure, however, as a consequence of the high degree of polymerization, the attractive forces between the individual molecules of the structural elements are to a large extent caused by primary bonds.

Such systems owe their external appearance to a structure resulting from a more or less random

distribution of structural units. Owing to the lack of symmetry, the presence of flaws can be assumed to exist as pointed out by Griffith.⁴ These flaws play the same role as the holes in a liquid according to Eyring's theory,⁵ and represent the space into which the structural elements can slip, when the system is subjected to a stress. As a result of the smallness of the flaws relative to the volume of the structural elements, the mechanism of the movement of the structural elements is complicated. It may consist of a shear movement of the structural units along certain planes or of a migration of molecular segments throughout the structural unit, starting at one side of the unit and crossing it like a wave in the requisite direction.⁶ In any case, the individual unit cannot flow in any direction as the molecules in a liquid but moves only in directions which are made available by the location of the flaws.

This process of flow also requires a rotation of the structural units with their long axis in the direction of flow. In view of their more or less random distribution, the structural elements are at different angles to the general direction of flow. It happens, therefore, that units of different orientation tend to move in slightly different directions or even in the opposite direction in the

¹ Charles Mack, *J. App. Phys.* 17, 1086 (1946).

² T. S. Moore and T. F. Winmill, *J. Chem. Soc.* 101, 1635 (1912).

³ P. Pfeiffer, *Ann.* 398, 137 (1913).

⁴ A. A. Griffith, *Phil. Trans. Roy. Soc. London* 221, 163 (1921).

⁵ H. Eyring, *J. App. Phys.* 4, 283 (1936).

⁶ W. Kauzmann and H. Eyring, *J. Am. Chem. Soc.* 62, 3275 (1940).

extreme case. With the change in orientation of the structural units, the shape of the flaws is also altered, and at the sharp end of a flaw there is an excessive strain which results in a concentration of stress at that point accompanied by a reduction in stress at other places. For an elliptical flaw (with the long and short half axis of a and b) under the influence of an exterior stress S perpendicular to the long axis a , Griffith⁴ calculated the stress at the sharp edges to be equal to $S(2a+b)/b$. The motion of the structural units will become arrested at such points of stress concentration. As the process continues more units will cease moving, until the total internal stress becomes equal to the external stress, and the deformation of the plastic body under test comes to a stop. An increase in deformation can then only be obtained by the application of a higher stress. It is usually found that when the substance is unloaded and loaded again, deformation does not begin at the same stress as originally observed but at a higher value. This value is often found to be near the value of the stress applied in the experiment, before the unloading took place. Thus the deformation has increased the strength of the material, i.e., the stress which can be supported, and this phenomenon is called "work-hardening." As the number of structural units arrested decreases with time, a constant stress produces a continuously increasing deformation, the rate of which decreases with time. This type of deformation as a function of time is generally termed "creep." During this deformation the neighboring structural units, although moving in slightly different directions, must maintain contact at all points, or a crack will form and spread through the whole substance. To ensure contact, some distortion of the structure of the units must occur particularly at the boundaries. If this distortion becomes large enough in extent, so that the structural units are forced to move mainly in one direction, the creep will change into flow, i.e., the rate of deformation becomes constant.

The concept of creep can also be applied to thixotropic systems. A stress causes a continuous breakdown of the three-dimensional structures without permitting the formation of new connections between the structural elements. These changes in structure reveal themselves in the fact that as soon as the first plastic deformation is

obtained, further deformation requires a lower stress than the original. As a result of this, the strength, which can be supported by a thixotropic system, decreases with increasing deformation, and this process may be considered as a case of "work-softening."

CREEP AND WORK-HARDENING

The phenomenon of creep accompanied by work-hardening under constant stress can be represented by the same model, consisting of an elastic element and a number of viscous elements, as given for plastic flow in Part I.¹ Whereas in the latter case the deformability of the viscous elements is unlimited, the extensibility of the elements of a substance subject to creep must be assumed to be limited. A stress applied to such a system sets the pistons of all viscous elements in motion with yield values smaller than the applied stress. As the deformation proceeds, some of the elements reach their maximum extensibility resulting in a decrease in the number of elements deformed and in a reduction of the total strain rate. Finally the motion of all pistons of the viscous elements comes to a stop, and the strain rate becomes zero. With n elements having reached the limit of their extensibility, the original yield value is increased to that of the $n+1$ st element, and it requires a stress increment to bring about further deformation. This condition holds only, if the direction of the load stress is kept constant. If a tensile stress is applied to the system for some time and then the direction of loading reversed from tension to compression, the pistons of all elements which have reached their maximum extensibility will start moving again in the opposite direction, and the total strain rate of the system will be increased. This behavior explains the observation that metals, after they are plastically deformed, undergo comparatively large plastic deformations by small loads, when the direction of loading is reversed.⁷

The strain rate of the system can be expressed by Eq. (5) given in Part I

$$v = \sum (S - S_f) / \eta = n \varphi(S) / \eta, \quad (1)$$

where the stress function $\varphi(S)$ is either bS or bS_0 depending upon the applicability of Eq. (3) or

⁷A. Nadai, *Plasticity* (McGraw-Hill Book Company, Inc., New York, 1931).

(20) given in Part I to the system under test. For plastic flow n , the number of viscous elements deformed, is constant at constant stress. In the case of creep, n decreases continually from a maximum to a minimum or to zero and can be a function of the strain or of time. For a system subject to work-hardening, the stress can therefore be a function of strain and strain rate (strain-hardening) or a function of time and strain-rate (time-hardening). Since the time is always a function of the strain for this system, the stress function may be written in the following general form:

$$S = f(\epsilon, v). \quad (2)$$

Differentiation with respect to time gives

$$\frac{dS}{dt} = \frac{\partial S}{\partial \epsilon} \frac{d\epsilon}{dt} + \frac{\partial S}{\partial v} \frac{dv}{dt}. \quad (3)$$

$\partial S / \partial \epsilon$ measures the rate of change of stress with plastic strain at constant strain rate, and may be compared to a coefficient of work-hardening, ψ . In a similar way $\partial S / \partial v$ measures the rate of change of stress with changing strain rate at constant strain. This term has the dimensions of a coefficient of viscosity and may be termed coefficient of plasticity, ϕ . If the deformation is carried out at constant stress, dS is zero, and Eq. (3) becomes

$$\psi d\epsilon / dt = \psi v = -\phi dv / dt. \quad (4)$$

ϕ is a stress function divided by the strain rate, $\phi = \varphi(S) / v$, and

$$-dv / dt = v^2 \psi / \varphi(S), \quad (5)$$

i.e., the decrease in strain rate is proportional to the square of the strain rate and to $\psi / \varphi(S)$. At constant stress $\varphi(S)$ is constant, and the coefficient of work-hardening, ψ , can be constant or variable. Equation (5) also implies that, when $\psi = 0$, then $dv = 0$, and v becomes constant which is the condition for plastic flow.

(a) Coefficient of Work-Hardening is Constant at Constant Stress and Time-Hardening

With ψ being constant, the term $\psi / \varphi(S)$ also becomes constant. If it is replaced by $1/k$, then

$$-v^{-2} dv / dt = 1/k,$$

and on integration between v and v_1 , the strain

rate obtained 1 second after application of the constant stress, and t and 1 second

$$v = d\epsilon / dt = k / (t - 1 + k/v_1). \quad (6)$$

On second integration between ϵ and 0, t and 0

$$\epsilon = k \ln (t + k/v_1). \quad (7)$$

At $t = 0$, $\epsilon = 0$, hence $k/v_1 = 1$ and $k = v_1$, and

$$\epsilon = k \ln (t + 1). \quad (8)$$

In this equation and the ones to follow, ϵ represents the plastic strain only. With systems of sufficiently large elastic deformation, ϵ can be replaced by the difference between the total and elastic strains.

With $k = v_1$, it follows from Eq. (6) that

$$v = k/t. \quad (9)$$

Introducing this term into Eq. (1) for v and $v_1 = v$ at $t_1 = 1$ second gives on division of both equations for the condition of constant stress:

$$v/v_1 = t_1/t = n/n_1.$$

This term indicates that, since $n < n_1$, the number of elements under deformation decreases with time, and the creep is caused by time-hardening.

It follows from Eq. (8) that the strain is a linear function of the logarithm of time, but since the proportionality factor is constant only for a constant stress, k varies with the stress.

For the evaluation of the coefficient of work-hardening, ψ , it is necessary to define ϕ and $\varphi(S)$. If Eq. (3) of Part I, $S/S_0 = (v/v_0)^b$ is applied then $\psi = \partial S / \partial v = bS/v$, and $\varphi(S) = bS$, therefore since $k = \varphi(S) / \psi$

$$\psi = bS/k, \quad (10)$$

$$\phi = bSt/k. \quad (11)$$

These relationships give a coefficient of work-hardening which is constant at constant stress, and a coefficient of plasticity, which increases linearly with time.

Substituting the above terms for ψ and ϕ in Eq. (3) and replacing k by vt gives

$$dS/S = bdt/t + b dv/v,$$

which on integration between S and S_0 , t and t_0 , v and v_0 yields

$$S/S_0 = (vt/v_0 t_0)^b, \quad (12)$$

*

confirming that creep is caused by time-hardening. The constant term of viscosity for this relationship is obtained from Eq. (1) as follows:

$$\eta = nbS/v = n_0bS_0/v_0.$$

$$\text{Since } v/v_0 = (S/S_0)^{1/b}t_0/t$$

$$n/n_0 = (S/S_0)^{(1-b)/b}t_0/t$$

and the expression for viscosity becomes

$$\eta = n_0b(S/S_0)^{(1-b)/b}(t_0/t)S/v, \quad (13)$$

in which S and v are the only vectors. For this case the distribution of yield values is such that their sum is equal to $n_0(1-b)(S/S_0)^{(1-b)/b}t_0S/t$. This equation in combination with Eq. (15) given in Part I leads to the following term for the relaxation of stress at constant strain:

$$-dS/dt = Ev = n_0b(S/S_0)^{(1-b)/b}(t_0/t)SE/\eta,$$

and on integration between S and S_0 , the stress at the moment where the strain is kept constant, t and 0

$$(S_0/S)^{(1-b)/b} = n_0(1-b)(S_0/S_0)(1-b)/b \text{ to } (E/\eta) \ln(t+1) + 1. \quad (14)$$

Since $n_0t_0(S_0/S_0)^{(1-b)/b} = n_0t_0$, where n_0 and t_0 are the number of elements deformed and the time at the moment where the strain is kept constant, and $E/\eta = 1/\tau$, where τ is the relaxation time, this equation becomes

$$(S_0/S)^{(1-b)/b} = (1-b)n_0t_0 \ln(t+1)/\tau + 1. \quad (15)$$

Although this equation has been obtained for a system having one mechanism of elasticity and viscosity, it is also applicable to systems with several such mechanisms, in which case $1/\tau = \sum 1/\eta / \sum 1/E$, as outlined in Part I. The term $(1-b)n_0t_0/\tau$ varies with the number of elements under deformation at the moment where the strain is kept constant, and consequently varies not only with the initial stress, S_0 , but also with the time t_0 during which the substance was deformed, before the strain was kept constant. Replacing this term by $1/\ln t_r$, where t_r is the variable relaxation time, changes above equation into:

$$(S_0/S)^{(1-b)/b} = \ln(t+1)/\ln t_r + 1, \quad (16)$$

and $\ln t_r$ is the logarithm of the time at which the original stress is reduced to $S_0/2^{b/(1-b)}$. This equation

as well as the others to follow in connection with work-hardening holds only when the system is still subject to work-hardening during the relaxation of stress. Consequently the stress relaxes to a certain lower limit which can be close to the value of the initial stress. In this case the substance remains under strain, and further stress relaxation can be only obtained by heating the substance, which process is carried out by the annealing of glass and tempering of steel.

Applying the Prandtl Eq. (20) given in Part I, $\exp[(S-S_0)/S_0] = (v/v_0)^b$, to the case of time-hardening gives

$$\exp[(S-S_0)/S_0] = (vt/v_0t_0)^b, \quad (17)$$

the coefficient of viscosity is

$$\eta = n_0b \exp[(S-S_0)/bS_0](t_0/t)S_0/v \quad (18)$$

and the stress at constant strain relaxes according to

$$\exp[(S_0-S)/bS_0] = \ln(t+1)/\ln t_r + 1, \quad (19)$$

$\ln t_r$ being the logarithm of the time at which the initial stress is reduced to a value of $S_0 - bS_0 \ln 2$.

Equations (12) and (17) express the stress as a function of strain rate and time, $S = f(v, t)$. Differentiating this term gives the following general expression:

$$dS = \frac{\partial S}{\partial t}dt + \frac{\partial S}{\partial v}dv = \theta dt + \phi dv, \quad (20)$$

which implies that at constant stress $\partial S/\partial t = \theta$ approaches zero with high values of t . In this case dv also approaches zero, i.e., at deformations of long duration under constant stress the strain rate becomes finally constant, and creep changes into plastic flow.

These equations have been obtained by assuming, ψ , the coefficient of work-hardening or strain-hardening to be constant at constant stress. Equation (20) leads to an interesting solution when the coefficient of time-hardening is assumed to be constant, at constant stress, in which case

$$\theta dt + \phi dv = \theta dt + \varphi(S)dv/v$$

and

$$\theta/\varphi(S) = 1/\alpha = -v^{-1}dv/dt$$

where α is a constant with the dimension of t .

Integration between v and v_1 , t , and 0 gives

$$v = d\epsilon/dt = v_1 \exp(-t/\alpha), \quad (21)$$

and on second integration between ϵ and 0, t and 0

$$\epsilon = \alpha v_1 [1 - \exp(-t/\alpha)]. \quad (22)$$

The term αv_1 has the dimension of ϵ and is obviously the limiting value of ϵ , at which creep due to time-hardening at constant stress changes into plastic flow, and is equal to $\epsilon_{\max}$.

With $\varphi(S) = bS$ and assuming that α is a constant independent of the stress, the following equation for the stress as a function of strain rate and time is obtained:

$$S/S_0 = (v/v_0)^b \exp[b(t-t_0)/\alpha], \quad (23)$$

from which results the following expression for the coefficient of viscosity

$$\eta = n_0 b (S/S_0)^{(1-b)/b} \exp[-(t-t_0)/\alpha] S/v. \quad (24)$$

The corresponding equations for the condition $\varphi(S) = bS_0$ are:

$$\exp[(S-S_0)/S_0] = (v/v_0)^b \exp[b(t-t_0)/\alpha], \quad (25)$$

$$\eta = n_0 b \exp\{[\alpha(S-S_0) - bS_0(t-t_0)]/\alpha b S_0\} S_0/v. \quad (26)$$

Another solution of Eq. (20) is obtained if $\theta = \varphi(S)/at$, where "a" is a constant, in which case at constant stress

$$\theta dt + \phi dv = \varphi(S) dt/at + \varphi(S) dv/v,$$

and

$$dt/at = -dv/v.$$

With $\varphi(S) = bS$ and assuming that a is a constant independent of the stress and equal to b/c , where c is a new constant, the following equations are obtained: for the stress as a function of strain rate and time

$$S/S_0 = (v/v_0)^b (t/t_0)^c \quad (27)$$

for the coefficient of viscosity

$$\eta = n_0 b (S/S_0)^{(1-b)/b} (t_0/t)^{c/b} S/v \quad (28)$$

and for the relaxation of stress at constant strain

$$(S_t/S)^{(1-b)/b} = (t/t_r)^{(b-c)/b} + 1. \quad (29)$$

(b) Coefficient of Work-Hardening is Variable at Constant Stress and Strain-Hardening

In the previous section the case has been treated for the coefficient of work-hardening, ψ , being constant at constant stress, hence $\psi/\varphi(S) = \text{constant}$. If this term is a function of the strain, so that $\psi/\varphi(S) = 1/\epsilon$, Eq. (5) takes the following form for the condition of constant stress

$$-v^{-2} dv/dt = \psi/\varphi(S) = 1/\epsilon. \quad (30)$$

Substituting the term for ψ in Eq. (3) gives for $\varphi(S) = bS$

$$S/S_0 = (v\epsilon/v_0\epsilon_0)^b, \quad (31)$$

expressing the stress as a function of the strain and the strain rate, and the coefficient of work-hardening becomes a coefficient of strain-hardening. The difference between time-hardening and strain-hardening will be discussed in detail in a later section. It may be mentioned here that the coefficient of strain-hardening obtained from Eq. (30) $\psi = bS/\epsilon$, decreases considerably less with increasing deformation than the coefficient of time-hardening obtained from Eq. (13), $\theta = bS/t$, for the same value of bS . The latter term approaches zero with high values of t , which limit is only reached in the case of strain-hardening with very large values for ϵ , i.e., for elongations which correspond to a large multiple of the original length of the test piece.

The coefficient of viscosity for a substance subject to strain-hardening according to above equation is

$$\eta = n_0 b (S/S_0)^{(1-b)/b} (\epsilon_0/\epsilon) S/v. \quad (32)$$

For the evaluation of the relaxation of stress at constant total strain, the plastic strain is equal to the elastic strain but of opposite sign, and can therefore be replaced by $-S/E$. Introducing this term into Eq. (32) gives the stress relaxation as follows:

$$(S_t/S)^{(1-b)/b} = t/t_r + 1, \quad (33)$$

where $1/t_r = n_0(1-2b)/\tau$ and t_r depends upon the value of S , and ϵ_0 , the stress and strain prior to the moment where the strain is kept constant.

Introducing into Eq. (31) the term $\varphi(S) = bS_0$ leads to

$$\exp[(S-S_0)/S_0] = (v\epsilon/v_0\epsilon_0)^b \quad (34)$$

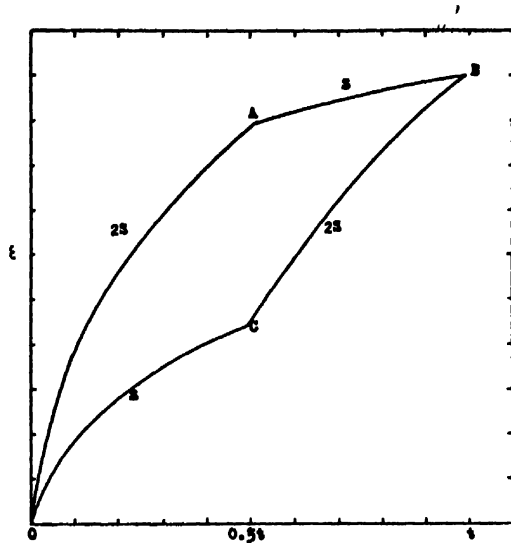


FIG. 1. Effect of strain-hardening: $S/S_0 = (v/v_0)^b$; $b = 0.5$.

and for the stress relaxation at constant strain

$$\left[\frac{(S - bS_0)}{S_0} \right] \exp \left[\frac{(S_0 - S)}{bS_0} \right] = n\dot{\epsilon}/\tau + (S_0 - bS_0)/S_0, \quad (35)$$

which is not a simple expression for the stress relaxation as a function of time. A better solution is obtained by substituting the strain in equation (34) by its time function as follows:

$$\exp \left[\frac{(S - S_0)}{S_0} \right] = (v/v_0)^{2b} (t/t_0)^b \quad (36)$$

which gives the coefficient of viscosity as

$$\eta = n_0 b (S_0/v) (t_0/t)^{1/b} \exp \left[\frac{(S - S_0)}{2bS_0} \right] \quad (37)$$

and the following stress relaxation as a function of time

$$\exp \left[\frac{(S_0 - S)}{2bS_0} \right] = (t/t_r)^{1/b} + 1 \quad (38)$$

where $1/t_r^{1/b} = n\dot{\epsilon}_r^{1/b}$.

In analogy with Eq. (27), ψ can also be assumed to be equal to $\varphi(S)/b\epsilon/m$, where m is a new constant, in which case the stress as a function of strain and strain rate for $\varphi(S) = bS$ takes the form

$$S/S_0 = (v/v_0)^b (\epsilon/\epsilon_0)^m. \quad (39)$$

(c) Work-Hardening Depending upon Strain and Time

The equations given in the previous sections express the stress as functions of strain rate and time or as functions of strain rate and strain. Cases may occur where the stress is a function of

strain, time and strain rate according to

$$S = f(\epsilon, t, v)$$

and

$$\frac{\partial S}{\partial \epsilon} d\epsilon + \frac{\partial S}{\partial t} dt + \frac{\partial S}{\partial v} dv = \psi d\epsilon + \theta dt + \phi dv. \quad (40)$$

In these cases the number of elements under deformation at constant stress decrease with increasing strain and time.

With $\phi = bS/v$, the application of the corresponding terms for ψ and θ previously given results in the following possible solutions of this equation:

$$S/S_0 = (v\epsilon t/v_0\epsilon_0 t_0)^b \quad (41)$$

$$S/S_0 = (vt/v_0 t_0)^b (\epsilon/\epsilon_0)^m \quad (42)$$

$$S/S_0 = (v\epsilon/v_0\epsilon_0)^b \exp [b(t-t_0)/\alpha] \quad (43)$$

$$S/S_0 = (v\epsilon/v_0\epsilon_0)^b (t/t_0)^c \quad (44)$$

$$S/S_0 = (v/v_0)^b (t/t_0)^c (\epsilon/\epsilon_0)^m. \quad (45)$$

(d) Difference between Time-Hardening and Strain-Hardening

The equations previously given express the stress as functions of strain rate and time, as functions of strain rate and strain or both. In each case the strain is a function of time, and the strain and time functions can be mutually rearranged. The equations given in connection with time-hardening can be expressed as follows:

$$S/S_0 = (vt/v_0 t_0)^b = (v/v_0)^b \times \exp \{ b[\epsilon(S_0/S)^{1/b} - \epsilon_0]/v_0 t_0 \} \quad (46)$$

$$S/S_0 = (v/v_0)^b \exp [b(t-t_0)/\alpha] = \{ (v/v_0) + [\epsilon - \epsilon_0(S/S_0)^{1/b}]/\alpha v_0 \}^b \quad (47)$$

$$S/S_0 = (v/v_0)^b (t/t_0)^c = (v/v_0)^b -^c (\epsilon/\epsilon_0)^c. \quad (48)$$

Similar equations are obtained for the stress function $\exp [(S - S_0)/S_0]$.

It has been mentioned before that creep due to time-hardening finally changes into plastic flow. This change is borne out by Eqs. (46) and (47) and occurs as soon as the value of $\epsilon(S_0/S)^{1/b}$ in the exponential term has reached the value of ϵ_0 , or in Eq. (47) when $\epsilon_0(S/S_0)^{1/b}$ has the value of ϵ . Equation (48) which originally has been derived for creep due to time-hardening is similar

in form to Eq. (39) in connection with strain-hardening. The creep in this case is consequently due to strain-hardening and extends over a wide range of deformation when the values of c are smaller than 1.

The creep at constant stress is of interest, which in the case of Eq. (46) is a linear function of the logarithm of time, and its magnitude increases with increasing stress. In the case of the equations in connection with strain-hardening, the logarithm of strain is a linear function of the logarithm of time, and its slope is independent of the stress applied. For Eq. (47) the logarithm of $\epsilon_{\max}/(\epsilon_{\max}-\epsilon)$ is as shown before a linear function of time independent of the stress applied.

This difference in creep further reveals itself in the behavior of substances subject to work-hardening when they are submitted to stresses in succession. Suppose a test piece subject to strain-hardening is first stretched under a constant stress $2S$ until A in Fig. 1 is obtained at $\frac{1}{2}t$. At this point the stress is reduced to S and kept constant for a second time interval of $\frac{1}{2}t$ during which the creep follows curve AB . A second test piece is subjected to the reversed order of loading, i.e., is first stretched under a constant stress S for the time $\frac{1}{2}t$, and then the stress is increased to $2S$. In this case the creep first follows curve OC and finally meets the strain-time curve at point B and identical time t . Thus if a substance subject to strain-hardening is deformed under various stresses, the strain at a given time will be the same irrespective of the order of loading. This behavior, of course, applies only to the plastic strain and not to the total strain (=elastic + plastic strain). Strain-time diagrams similar to that of Fig. 1 are obtained for substances whose creep is connected with strain-hardening irrespective of the derivation of these equations.

A substance, which is subject to time-hardening in accordance with Eq. (46), when successively stretched under the constant stresses S and $2S$, gives a strain-time diagram as presented in Fig. 2. Here the strain obtained after a certain time depends upon the order of loading, and the curves do not meet at the same time t . The strain of such a substance obtained at a given time depends therefore upon previous operations.

Substances obeying Eq. (47) take an intermediate position, in that they show under stress

a change from creep to plastic flow as in the case of time-hardening and, when successively stretched, behave like substances due to strain-hardening.

For the cases where the stress is a function of strain-rate, strain and time, it can be shown in the same manner, that the creep is independent of the history of the substance and finally changes into creep due to strain-hardening only.

CREEP AND THIXOTROPY

It has been shown that creep in combination with work-hardening is a process which increases the stress which can be supported by a substance. Thixotropic systems also show the phenomenon of creep which is however accompanied by breakdown of the structure and results in a decrease in stress which can be supported. Thus creep of thixotropic systems is opposite in effect to creep due to work-hardening. Thixotropic systems can therefore be compared with the same model consisting of an elastic element and a number of viscous elements coupled in series. A stress applied to such a system increases the number of elements deformed with time, and the strain rate increases. In view of this effect Eq. (20) can be applied to thixotropic systems with a negative

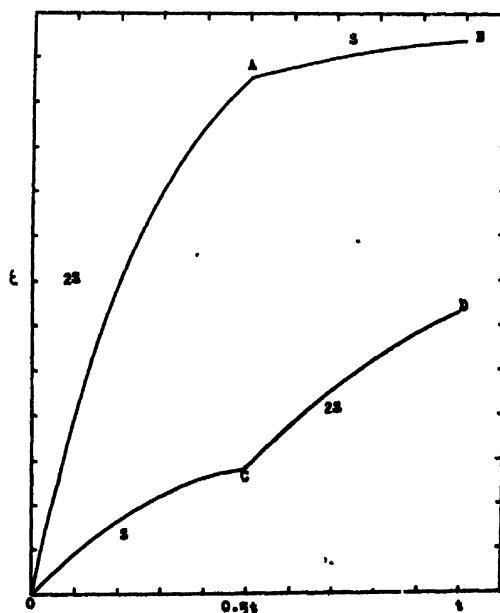


FIG. 2. Effect of time-hardening: $S/S_0 = (t_0/t_0 + \epsilon_0)^h$, $h = 0.5$.

coefficient of time-hardening:

$$dS = \frac{-\partial S}{\partial t} dt + \frac{\partial S}{\partial v} dv = -\theta dt + \phi dv. \quad (49)$$

This equation also implies that, if θ becomes zero at constant stress, then $dv=0$, and v becomes constant. This condition is fulfilled after breakdown of the thixotropic structure, the amount of which may depend upon the magnitude of the stress applied.

With $\theta = bS/t$, and $\phi = bS/v$, integration of Eq. (49) gives

$$S/S_0 = (v/v_0)^b (t/t_0)^{-b} = (vt_0/v_0t)^b \quad (50)$$

and with $\theta = cS/t$

$$S/S_0 = (v/v_0)^b (t_0/t)^c. \quad (51)$$

For both cases, the logarithm of strain is a linear function of the logarithm of time at constant stress, and the slopes are independent of the stress applied.

The viscosities resulting from these expressions are

$$\eta = \eta_0 b (S/S_0)^{(1-b)/b} (t/t_0) S/v, \quad (52)$$

$$\eta = \eta_0 b (S/S_0)^{(1-b)/b} (t/t_0)^{c/b} S/v. \quad (53)$$

With the thixotropic structure still subject to breakdown during relaxation at constant strain,

these equations lead to

$$(S_i/S)^{(1-b)/b} = (t/t_r)^2 + 1, \quad (54)$$

where $1/t_r^2 = [(1-b)/2]n_s/\tau t_s$, and

$$(S_i/S)^{(1-b)/b} = (t/t_r)^{(b+c)/b} + 1, \quad (55)$$

where $1/t_r^{(b+c)/b} = b(1-b)n_s/(b+c)\tau t_s^{c/b}$. In both cases the variable relaxation time t_r depends upon n_s , the number of elements deformed at the moment where the strain is kept constant and therefore varies with S , and t , the time during which the system was deformed prior to the beginning of relaxation. If the structure is completely destroyed, the stress relaxation follows Eq. (18) Part I, $(S_i/S)^{(1-b)/b} = t/t_r + 1$. Another possibility exists if the value of the stress falls below that of the attractive forces of the micelles for each other, in which case the thixotropic structure may be built up again during relaxation.

The expressions giving the stress as functions of strain rate, strain and/or time cover the creep of various systems. Many substances subject to work-hardening have mechanisms of deformation which also exhibit the property of elastic after-effect. Such systems will be discussed in a separate paper, and the presentation of data for substances subject to work-hardening and thixotropy will be reserved for this part.

Plastic Flow, Creep, and Stress Relaxation

Part III. Creep and Elastic After-Effect

CHARLES MACK

Technical and Research Department, Imperial Oil Limited, Sarnia, Ontario

(Received March 28, 1946)

A large number of substances show the phenomenon of elastic after-effect, and part of their deformation recovers on unloading as a function of time. This portion of the deformation at constant stress has a strain rate which decreases with time and is therefore comparable to creep. Expressions are given for the strain-time relationships of such systems, and the process of stress relaxation at constant strain is discussed. The equations given in connection with plastic flow, creep due to work-hardening, thixotropy, and creep in combination with elastic after-effect are applied to data given in the literature, and it is shown that these equations suffice to describe the deformation and relaxation mechanisms of a variety of materials such as metals, clay soil, food products, acrylic acid polymeride, polyvinyl chloride, cellulose acetate, manila ropes, paper laminates, phenolic molding compounds, rubber, asphalt, and bituminous pavements.

INTRODUCTION

THE theory of plastic flow and creep has been discussed in two previous papers.^{1,2} It has been shown that these mechanisms of deformation can be represented by a model consisting of an elastic element and a number of viscous elements all coupled in series. The viscous elements represent the structural units of the plastic system and have an unlimited extensibility under tension in the case of plastic flow. In the case of creep their extensibility is limited as a result of a resistance which is caused by the position of the structural units relative to each other. If elastic forces are the cause of limited extensibility, the structural units can be considered as being encased in an elastic sheath. A tensile stress acting on such an element causes a deformation which increases with time up to a certain limit. After removal of the load the elastic sheath will recover its original shape, but, since the element also possesses viscosity, the recovery is a function of time. This phenomenon is generally referred to as elastic after-effect. With structural units of this type present in a substance, the deformation mechanism can be described by models suggested by Burgers³ consisting of elastic and viscous elements coupled in series as well as in parallel.

Such a system is represented by Fig. 1, and the various elements are considered as being rigidly attached to each other. Elements *A* and *B* are perfectly elastic springs, and elements *C* and *D* consist of porous pistons moving in cylinders filled with a viscous liquid. When the model is subjected to a constant tensile stress, spring *A* will immediately elongate followed by movements of the pistons of the viscous elements *C* and *D*. The deformation of element *C* necessitates

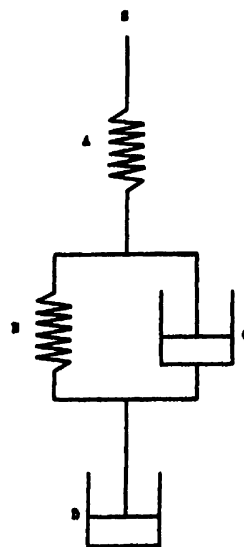


FIG. 1. Model of plastic deformation with elastic after-effect.

¹ Charles Mack, *J. App. Phys.* 17, 1086 (1946).

² Charles Mack, *J. App. Phys.* 17, 1093 (1946).

³ G. M. Burgers, "First report on viscosity and plasticity," *Acad. Sci. Amsterdam* (1935).

an elongation of spring *B* and part of the stress is taken over by this spring.

The deformation of elements *C* and *D* can be completely or partially simultaneous, or it can occur that element *D* will be deformed after element *C* has reached its maximum extensibility. The latter case is often met with rubber-like substances containing molecules of high degree of polymerization which are large in one dimension only. Such substances consist of a mass of entangled chain molecules which have a repeating unit. Any increase in length of such a substance corresponds to an increase in the distance between the ends of the individual chain molecules and to a change in their relative position. A tensile stress applied will first increase the distance between the individual links in the chain. This affects interatomic distances, and the deformation is purely elastic. Under continued stress the molecule will be further stretched by rotating about the C-C bonds of the main skeleton to a new set of positions. This deformation is also elastic, but requires a certain time, and corresponds to the deformation of the couple formed by elements *B* and *C*. Finally, whole portions of chains alter their relative position by sliding over each other, and this part of the deformation is wholly plastic.⁴

CREEP AND ELASTIC AFTER-EFFECT

(a) Deformation Mechanism of Model with Elastic After-Effect

The element causing the elastic after-effect consists of spring *B* coupled in parallel with the viscous element *C*. Here the strain is the same for each component. A tensile stress acting on this couple, however, is composed of the sum of the stresses for spring *B* and element *C*.

$$S = S_b + S_c = S_b + (S - S_b).$$

The strains of spring *A*, couple *BC*, and element *D* are additive,

$$\epsilon = \epsilon_a + \epsilon_b + \epsilon_d,$$

and substituting the strains by their time-functions leads to an expression for the total strain as functions of time at constant stress, which has been carried out by Burgers for a system containing purely viscous liquids in elements *C* and *D*.

The same procedure can be also applied to a system in which the viscous elements *C* and *D* are replaced by elements subject to plastic flow. With element *C* obeying the relationship $S/S_0 = (v/v_0)^b$, *S* must be replaced by $S_c = S - E\epsilon_b$, the stress acting on this element. Applying the corresponding term for the viscosity gives the strain of the couple *BC* as

$$\epsilon_b = \epsilon_{b \max} \{1 - [t/t_{or} + 1]^{-b/1-b}\}, \quad (1)$$

where $\epsilon_{b \max}$ is the maximum strain obtained under the stress applied and is S/E . The expression $1/t_{or}$ is equal to $n_0(1-b)E/\eta$ and has the dimension of inverse time. The term η/E is the same as the relaxation time (see Part I), but since its meaning is different, the term "orientation time," τ_{or} , has been proposed.⁴ The expression t_{or} varies with *S*. With element *C* obeying the Prandtl equation, $\exp[(S - S_0)/S_0] = (v/v_0)^b$, the strain as a function of time becomes

$$\exp(\epsilon_b/\epsilon_{b \max}) = (t/t_{or} + 1)^b, \quad (2)$$

where $1/t_{or} = n_0 E/\eta$.

Equations (1) and (2) show that the strain-time relationships are the same for loading and unloading but of opposite sign. This behavior is of importance, since by measuring the total strain as a function of time during loading and observing the recovery after unloading as a function of time, the strains of the individual elements at a given time can be obtained for the condition of loading.

In the case where element *C* is supposed to be subject to work-hardening, the latter effect influences the deformation to such an extent that soon after the application of a constant tensile stress the elongation of element *C* comes to a stop. The deformation obtained is probably so small that it has no practical significance.

The system causing the elastic after-effect need not be restricted to one spring and viscous or plastic element coupled in parallel and may be composed of several springs and elements all coupled in parallel or may consist of a series of one spring and one element coupled in parallel and the whole series coupled in series. Such systems have been discussed by Alfrey and Doty.⁵

⁴ R. F. Tuckett, *Chem. and Ind.* 62, 430 (1943).

⁵ Alfrey and Doty, *J. App. Phys.* 16, 700 (1945).

TABLE I. Strain as function of time at constant stress.

Basic relationships for stress as functions of strain rate	Strain as function of time at constant stress	
	for element D	for spring B after unloading
(1) $S/S_0 = v/v_0$	$\epsilon = \text{const. } t$	$\epsilon_u/\epsilon = \exp(t/\tau_{or})$
(2) $S/S_0 = (v/v_0)^b$	$\epsilon = \text{const. } t$	$\epsilon_u/\epsilon = [(t/t_{or}) + 1]^{1/(1-b)}$
(3) $\exp(S - S_0)/S_0 = (v/v_0)^b$	$\epsilon = \text{const. } t$	$\epsilon_u - \epsilon = \epsilon_u b \ln[(t/t_{or}) + 1]$
(4) $S/S_0 = (vt/v_0 t_0)^b$	$\epsilon = \text{const. } \ln(t+1)$	
(5) $\exp[(S - S_0)/S_0] = (vt/v_0 t_0)^b$	$\epsilon = \text{const. } \ln(t+1)$	
(6) $S/S_0 = (v/v_0)^b \exp[b(t - t_0)/\alpha]$	$\epsilon = \text{const. } [1 - \exp(-t/\alpha)]$	
(7) $\exp[(S - S_0)/S_0] = (v/v_0)^b \exp[b(t - t_0)/\alpha]$	$\epsilon = \text{const. } [1 - \exp(-t/\alpha)]$	
(8) $S/S_0 = (v/v_0)^b (t/t_0)^c$	$\epsilon = \text{const. } t^{(b+c)/b}$	
(9) $\exp[(S - S_0)/S_0] = (v/v_0)^b (t/t_0)^c$	$\epsilon = \text{const. } t^{(b+c)/b}$	
(10) $S/S_0 = (ve/v_0 e_0)^b$	$\epsilon = \text{const. } t^b$	
(11) $\exp[(S - S_0)/S_0] = (ve/v_0 e_0)^b$	$\epsilon = \text{const. } t^b$	
(12) $S/S_0 = (v/v_0)^b (\epsilon/\epsilon_0)^m$	$\epsilon = \text{const. } t^{b/(b+m)}$	
(13) $\exp[(S - S_0)/S_0] = (v/v_0)^b (\epsilon/\epsilon_0)^m$	$\epsilon = \text{const. } t^{(b+m)/b}$	
(14) $S/S_0 = (v/v_0)^b (t_0/t)^c$	$\epsilon = \text{const. } t^{(b+c)/b}$	

Where const. has different meaning for each equation.

(b) Stress Relaxation of Systems with Elastic After-Effect

If the system is strained under tension, and then the strain kept constant, the stress relaxation can consist of two different processes depending upon the value of the moduli of elasticity of the elastic springs A and B relative to each other. When E_a of spring A is larger than E_b of spring B , spring A will contract accompanied by an elongation of the couple B, C and element D . When E_b is larger than E_a , the couple B, C will contract and spring A and element D will be under tension. In either case the relaxation of stress with time follows the general equation:

$$-dS/E_a dt = v = v_0 + v_d. \quad (3)$$

For the case of elements C and D containing perfect liquids with coefficients of viscosity η_a and η_d , this equation becomes on integration

$$S = M \exp[-(p+q)t] + N \exp[-(p-q)t], \quad (4)$$

where

$$p = \frac{1}{2} \left(\frac{E_a}{\eta_d} + \frac{E_a + E_b}{\eta_a} \right),$$

$$q = \left[\frac{1}{4} \left(\frac{E_a}{\eta_d} + \frac{E_a + E_b}{\eta_r} \right)^2 - \frac{E_a E_b}{\eta_r \eta_d} \right]^{1/2},$$

and M and N are constants with the dimensions

of stress. It will be noticed that the first exponential term dies away more rapidly than the second and finally approaches the value of zero. At this point the couple B, C is at its maximum elongation for the particular stress at that moment, if E_a is larger than E_b , and the relaxation proceeds according to the second exponential term. Thus the logarithm of stress is first curvilinear with time and becomes finally linear.

This relaxation mechanism is also met with systems for which the elements C and D are subject to plastic flow or creep. In this case Eq. (3) is also applicable. Substituting v_d from $S/S_0 = (v/v_0)^b$, and integrating between S and S_1 , and t_1 , S , and t , being the stress and time at which S becomes equal to S_1 , gives

$$(S_1/S)^{(1-b)/b} = [(1-b)/bt_1](t-t_1) + 1. \quad (5)$$

The relaxation process is similar for systems with elements C and D having flow properties different from those of the above case. The corresponding terms are summarized together with the strain/time relationships of the various elements in Tables I and II. It will be noticed that the equations for the final stage of stress relaxation for systems with elastic after-effect are similar in the cases of Eqs. (3), (9), (11), and (13) given in Table II, and do not always allow for a differentiation between relaxation due to plastic flow or due to creep based on strain-hardening.

TABLE II Relaxation of stress at constant strain.

Basic relationships for stress-rate functions of strain-rate	Stress relaxation for systems free of elastic after effect	Final stage of stress relaxation for systems with elastic after effect
(1) $S/S_0 = v/v_0$	$S/S_0 = \exp(t/\tau)$	$S/S_0 = \exp(t-t_0)/\tau$
(2) $S/S_0 = (v/v_0)^b$	$(S_0/S)^{(1-b)} = (t/t_0) + 1$	$(S_0/S)^{(1-b)} = (t-t_0)/t_0 + 1$
(3) $\exp[(S-S_0)/S_0] = (v/v_0)^b$	$\exp[(S_0-S)/bS_0] = (t/t_0) + 1$	$\exp[(S_0-S)/bS_0] = (t-t_0)/t_0 + 1$
(4) $S/S_0 = (v/v_0)^{b_1}$	$(S_0/S)^{(1-b)} = (\ln t/\ln t_0) + 1$	$(S_0/S)^{(1-b)} = (\ln t - \ln t_0)/\ln t_0 + 1$
(5) $\exp[(S-S_0)/S_0] = (v/v_0)^{b_1}$	$\exp[(S_0-S)/bS_0] = (\ln t/\ln t_0) + 1$	$\exp[(S_0-S)/bS_0] = (\ln t - \ln t_0)/\ln t_0 + 1$
(6) $S/S_0 = (v/v_0)^b \exp[b(t-t_0)/\alpha]$	$(S_0/S)^{(1-b)} = (\alpha/t_0)[1 - \exp(-t/\alpha)] + 1$	$(S_0/S)^{(1-b)} = (\alpha/t_0)[1 - \exp[-(t-t_0)/\alpha]] + 1$
(7) $\exp[(S-S_0)/S_0] = (v/v_0)^b \exp[b(t-t_0)/\alpha]$	$\exp[(S_0-S)/bS_0] = (\alpha/t_0)[1 - \exp(-t/\alpha)] + 1$	$\exp[(S_0-S)/bS_0] = (\alpha/t_0)[1 - \exp[-(t-t_0)/\alpha]] + 1$
(8) $S/S_0 = (v/v_0)^b(t/t_0)^c$	$(S_0/S)^{(1-b)} = (t/t_0)^{c-1} + 1$	$(S_0/S)^{(1-b)} = [(t-t_0)/t_0]^{c-1} + 1$
(9) $\exp[(S-S_0)/S_0] = (v/v_0)^b(t/t_0)^c$	$\exp[(S_0-S)/bS_0] = (t/t_0)^{c-1} + 1$	$\exp[(S_0-S)/bS_0] = (t-t_0)/t_0 + 1$
(10) $S/S_0 = (v/v_0)^{b_1}$	$(S_0/S)^{(1-b)} = (t/t_0) + 1$	$(S_0/S)^{(1-b)} = (t-t_0)/t_0 + 1$
(11) $\exp[(S-S_0)/S_0] = (v/v_0)^{b_1}$	$\exp[(S_0-S)/2bS_0] = (t/t_0) + 1$	$\exp[(S_0-S)/2bS_0] = [(t-t_0)/t_0] + 1$
(12) $S/S_0 = (v/v_0)^b(e/e_0)^m$	$(S_0/S)^{(1-b)} = (t/t_0)^{b/(b+m)} + 1$	$(S_0/S)^{(1-b)} = [(t-t_0)/t_0]^{b/(b+m)} + 1$
(13) $\exp[(S-S_0)/S_0] = (v/v_0)^b(e/e_0)^m$	$\exp[(S_0-S)/(b+m)S_0] = (t/t_0)^{b/(b+m)} + 1$	$\exp[(S_0-S)/(b+m)S_0] = [(t-t_0)/t_0]^{b/(b+m)} + 1$
(14) $S/S_0 = (v/v_0)^b(t_0/t)^c$	$(S_0/S)^{(1-b)} = (t/t_0)^{c-1} + 1$	$(S_0/S)^{(1-b)} = (t-t_0)/t_0 + 1$

Where t_0 has different meaning for each equation

COMPARISON OF EQUATIONS FOR PLASTIC FLOW AND CREEP WITH DATA FROM LITERATURE

There is a multitude of strain/time data presented in the literature, but they are in most cases restricted to the condition of loading only, and a possible recovery of strain as a function of time is seldom taken into consideration. It is therefore difficult to obtain a complete picture of the flow characteristics of the substance under test, and only a few cases presented in the literature may be discussed.

(a) Plastic Flow

Examples for plastic flow free of elastic after-effect according to Eqs. (2) and (3) of Table I have been quoted in a previous paper.¹

(b) Creep and Elastic After-Effect

Leaderman⁶ measured the creep of polyvinyl chloride plasticized with tricresyl phosphate as a function of time under a constant stress and observed the recovery of strain after unloading. The sample showed almost complete recovery and corresponded essentially to the model presented in Fig. 1 minus element *D*. The curves given indicate a linear relationship between the strain and the logarithm of time and are in agreement with Eq. (3) in Table I for the couple *B*, *C*.

Lutz and Himmelfarb⁷ measured the creep of manila ropes under constant tensile stress. The curves given showed for the largest part a linear relationship between strain and logarithm of time, the slope of which increased with increasing stress. No observation of the recovery was made, however, it can be assumed from the fibrous structure of the material, that the manila ropes conform to the same equation as the plasticized polyvinyl chloride.

(c) Creep and Time-Hardening

McVetty⁸ suggested the following relationship between strain and time at constant stress for steel and its alloys and copper at elevated

⁶ Herbert Leaderman, Ind. Eng. Chem. **35**, 314 (1943).

⁷ Lutz and Himmelfarb, Proc. Am. Soc. Test. Mat. **40**, 125 (1940).

⁸ P. G. McVetty, Proc. Am. Soc. Test. Mat. **43**, 707 (1943).

temperature:

$$\epsilon = \epsilon_0 - (c/\alpha) \exp(-\alpha t) + v_0 t,$$

where ϵ is the total strain, ϵ_0 the strain at which the strain-rate becomes constant, $v_0 = \text{constant}$ strain-rate, $t = \text{time}$, and c and α are constants depending upon the material and temperature. It was suggested that the exponential term serves to define the strain-time relationship during creep. Evidently the term c/α has the dimension of ϵ and corresponds to ϵ_0 being the strain at which the exponential term approaches zero, hence

$$\epsilon = \epsilon_0 [1 - \exp(-\alpha t)] + v_0 t.$$

The first term of this equation can refer to the mechanical behavior of an elastic element coupled in parallel with a purely viscous element according to Eq. (1) in Table I, which is highly improbable in the case of metals, and therefore refers to creep due to time-hardening according to Eq. (7) in Table I with $1/\alpha$ equal to α in the above equation.

The relationship between stress and the constant strain rate v_0 was found to obey the following hyperbolic sine term

$$v_0 = v_1 \sinh(S/S_1) = \frac{1}{2} [\exp(S/S_1) - \exp(-S/S_1)]$$

in which the second exponential term approaches zero with high values of S , and this equation becomes similar to Eq. (3) in Table I.

The same equation also forms the basis for the stress relaxation of copper at constant strain, for which Boyd⁹ gave the following expression at elevated temperatures around 200°C for the last stages of relaxation

$$S = S_1 [1 - A \ln(Bt + 1)]$$

or

$$\exp[(S_1 - S)/AS_1] = Bt + 1,$$

which is the same as the relaxation term (3) in Table II with $A = b$ and $B = 1/t_r$. For the first stage the relationship between stress and logarithm of time forms a curve and seems to be influenced by the elastic after-effect as discussed in the previous section.

Bituminous pavements¹⁰ composed of 80 per-

cent of sand, 10 percent of silica dust, and 10 percent of asphalt were found to correspond in their deformation mechanism to the model given in Fig. 1. The elasticity of spring A of this model is mainly due to the small air pockets present through the pavement, and element D is subject to time-hardening according to Eq. (4) in Table I, $S/S_0 = (v/v_0 t_0)^b$. To illustrate the time-hardening effect, the data below may be given which were obtained at 15°C for such a mixture which was compacted in a mold of 5.08 cm in diameter to a cylinder 2.31 cm high under a pressure of 220 kg/cm². The cylinder having an air content of 6.73 percent by volume was subjected to compression tests under constant stress, in such a way that when the deformation came to a stop, i.e., when the initial yield value was increased to that of the stress, the load was increased by constant increments of 0.88 kg/cm². For the non-recoverable deformation the following values of $k = v$ were obtained:

Stress in kg/cm ²	Yield value in kg/cm ² at beginning of test	$k \times 10^3$
1.55	0.70	0.68
2.43	1.55	0.52
3.31	2.43	0.38
4.19	3.31	0.28
5.07	4.19	0.23
5.95	5.07	0.19
6.82	5.95	0.16
7.70	6.82	0.14
8.58	7.70	0.13
9.46	8.58	0.10

From the data the value of b is -0.89 , indicating that the increase in yield value is also affected by the stress as outlined in Part I of this paper.¹ During this stress range the density of 2.038 remained constant after the elastic recovery. At stresses in excess of 9.46 kg/cm² the values of k increased with increasing stress and the density dropped to 2.03 at the application of a stress of 12.10 kg/cm² resulting in opening of the structure and failure of the paving sample.

The modulus of elasticity of spring A was 3940 kg/cm² and that of spring B 9080 kg/cm². Measurements of the strain recovery after unloading showed element C to obey Eq. (2) in Table I, $S_b/S_0 = (v/v_0)^b$ with a value of b equal to 0.272. The relaxation of stress at constant strain was mainly due to the expansion of spring B , since its modulus of elasticity was higher than that of spring A .

⁹ John Boyd, Proc. Am. Soc. Test. Mat. 37, 218 (1937).

¹⁰ Charles Mack, unpublished data.

(d) Creep and Strain-Hardening

Creep data given by Sturm¹¹ for aluminum wire under tension at room temperature showed a linear relationship between the logarithm of strain and the logarithm of time with slopes which are practically independent of the stress applied. The stress was found to be a power function of the strain rate at a given deformation, hence the creep of aluminum wire corresponds to any of the Eqs. (8) to (13) in Table I and is caused by strain-hardening.

Scott Blair and Coppen¹² suggested the following equation for the relationship between stress, strain and time of plastic substances.

$$\psi = St^k/\epsilon,$$

where ψ is a constant measuring "firmness," and k a dimensionless constant. Both constants were found to be independent of the stress in some cases. In most cases ψ was found to vary with the stress, and the following equation was suggested, which was also independently proposed by Nutting.¹³

$$\psi_\beta = S^\beta t^k/\epsilon.$$

Expressing the stress as a function of strain rate and strain, changes these terms into

$$S/S_0 = (v/v_0)^k (\epsilon/\epsilon_0)^{1-k}$$

and

$$S/S_0 = (v/v_0)^{k/\beta} (\epsilon/\epsilon_0)^{(1-k)/\beta}.$$

Both equations cover creep in connection with strain-hardening and correspond in form to Eq. (12) in Table I. The authors found these relationships to hold for a variety of materials under compression at constant stress such as wet and dry clay, worked and unworked butter, cake, apple, cheese, potato, and acrylic acid polymeride. It appears that these substances are free of components giving rise to elastic after-effect or that the latter is so small as to be negligible.

Other substances such as plasticine and its mixtures with rubber and Vaseline showed partial recovery on unloading, and a linear relationship was found to hold between the logarithm of strain

and the logarithm of time. In view of the limitation in deformation of element *C* in Fig. 1 when subject to work-hardening, the elastic recovery appears to conform to the relationship (2) in Table I. The non-recoverable deformation of these materials corresponded to creep due to strain-hardening.

Findley¹⁴ found the following equation to hold for the creep of paper and canvas laminates, cellulose acetate sheets, and a series of phenolic molding compounds:

$$\epsilon - \epsilon_0 = mt^n,$$

where ϵ_0 is the strain at time $t=0$, and m and n are constants. Constant m was found to depend on the stress applied according to

$$m = C[\exp(kS) - 1],$$

where C and k are constants. Combining both terms, he arrived at the following equation for the strain rate

$$v = Cn^{n-1}[\exp(kS) - 1].$$

With high values of S , -1 can be neglected, which was introduced only in order to satisfy the condition, that $v=0$ when $S=0$; combining this equation with the corresponding term for v_0 gives after substituting bS_0 for $1/k$

$$\exp[(S-S_0)/S_0] = (v/v_0)^b (t/t_0)^{b(1-n)},$$

which is equivalent in form to Eq. (9) given in Table I in connection with strain-hardening. The cellulose acetate also showed the property of elastic after-effect at high stresses.

Buist and Seymour¹⁵ applied the Scott Blair equation mentioned before to data obtained from the compression and elongation of rubber and found a linear relationship between the logarithm of strain and logarithm of time at constant stress. Since the deformation of the couple *B, C* in Fig. 1 is very rapid in the case of rubber, the deformation as a function of time is measured mainly for element *D* and represents creep due to strain-hardening.

These results are in agreement with data given by Behre¹⁶ who measured the extrusion of rubber

¹¹ R. G. Sturm, Proc. Am. Soc. Test. Mat. **43**, 728 (1943).

¹² Scott Blair and J. Coppen, Soc. Chem. Ind. **60**, 190 (1941).

¹³ P. G. Nutting, Proc. Am. Soc. Test. Mat. **21**, 1162 (1921).

¹⁴ W. N. Findley, Modern Plastics **22**, 153 (Dec. 1944).
¹⁵ Buist and Seymour, Trans. Inst. Rub. Ind. [XVII] **2**, 366 (1944).

¹⁶ J. Behre, Rubber Chem. and Techn. **18**, 646 (1945) translated from Kautschuk **20**, 15 (1944).

as a function of time and found the following relationship to hold

$$V_{15} = K p^m,$$

where V_{15} is the volume of rubber extruded in 15 seconds, p = pressure applied, K and m are constants. Since the volume extruded and the pressure are proportional to the rate of shear and shearing stress, this term corresponds to Eq. (8) in Table I for strain-hardening with the time being constant.

(e) Creep and Thixotropy

An extensive investigation into the flow characteristics of a plastic asphalt was carried out by Saal and Labout,¹⁷ which consisted in the measurement of the deformation in a rotating cylinder viscometer as a function of time under constant shearing stress, and in measurements of the recovery after unloading and the stress relaxation at constant strain.

The logarithm of deformation was found to be proportional to the logarithm of time at constant stress for the first part with a proportionality factor of 0.38 independent of the stress. This relationship became then curvilinear and finally approached linearity again. The recovery-time curves, with deformation and time at the moment of unloading being 0, coincided with the first part of the deformation-time curves during loading. This part therefore corresponds to the deformation of the couple B, C , in Fig. 1. The deformation of element D was found to be proportional to a power function of time with a value of this

function higher than 1. This behavior is in agreement with the break-down of thixotropic structure according to Eq. (14) in Table I. Element C is free of thixotropic disturbance.

Re-evaluation of the stress/time curves during relaxation at constant deformation showed the logarithm of stress to be curvilinear with the logarithm of time for the initial stage due to the effect of the couple B, C . This relationship became then linear with a slope of -0.455 . Since the deformation of element C is proportional to $t^{b/(1-b)}$ according to Eq. (2) in Table I and $b/(1-b) = 0.38$, the stress relaxation as a function of time should have the same value for $b/(1-b)$ according to Eq. (2) in Table II. As this value is larger, it follows that the relaxation is still accompanied by break-down of the thixotropic structure and follows the Eq. (14) in Table II, $S_t/S = (t/t_0)^{(b+c)/(1-b)}$, with $b/(1-b) = 0.38$, b becomes 0.275 and c 0.055.

Values of c of this order are frequently found with asphalts. For example¹⁸ a thixotropic asphalt free of elastic disturbance showed in a coaxial falling cylinder viscometer under constant shearing stress at 15°C a linear relationship between the logarithm of deformation and logarithm of time for the period of structural break-down in agreement with Eq. (14) in Table I, $\epsilon = \text{const. } t^{(b+c)/b}$. With values of $(b+c)/b$ of 1.090 and 1.092 obtained at two different stresses and $b = 0.95$, c becomes 0.086.

These examples may suffice to show that the equations given have a wide field of application for the description of the deformation mechanism of plastic substances.

¹⁷ Saal and Labout, J. Phys. Chem. **44**, 149 (1940).

¹⁸ Charles Mack, unpublished data.

Asymptotic Solutions for the Normal Modes in the Theory of Microwave Propagation*

C. L. PEKERIS**

(Received July 2, 1946)

In this paper are presented several extensions of the "W.K.B." method in the asymptotic solution of differential equations. The developments were made in the course of a systematic study of the application of asymptotic solutions to the normal-mode theory of microwave propagation. In Section 1 a brief outline is given of the now standard normal-mode theory of propagation of microwaves in an atmosphere with a horizontally stratified index of refraction. Particular emphasis is given in this study to the case of a "surface duct." The case of "leaky modes" is studied in Section 2. Here the aim has been to obtain explicitly the terms after the leading one in the asymptotic expansion of the solution, in order to have an estimate of the order of magnitude of the error introduced by the use of the leading term only. This is achieved in Eq. (20), which is of general applicability to a wide class of physical problems. The first correction term to the phase-integral solution for the characteristic values of the normal modes is set out in Eqs. (28) to (30). An alternative asymptotic solution for the case of leaky modes, which includes the first correction terms, is given in Eqs. (30), (40), and (41). In Section 3 a direct method is given of determining the characteristic values from the phase-integral relation. If the modified index of refraction $y(h)$ is given by a power series (49) then with the aid of Eqs. (50)–(52) one can compute directly the characteristic value Λ_m by Eq. (46), using (47) and (48). The degree of success of this method is illustrated in Table I. The principal contribution of this study is in the development of asymptotic solutions for the case of *transitional modes* which are on the border line between the leaky modes and the trapped modes. The standard "W.K.B." method is inapplicable in this case. The results for the determination of the characteristic values and height-gain functions are as follows. One first obtains v_0 defined in (85), which in the case of an exponential model can be facilitated by the use of Fig. 2. The

quantities γ_m^0 , $d\gamma_m^0/dv_0$, and $[(\partial P(v_0)/\partial v) \cdot (\partial P(v_0)/\partial \gamma_m^0)]$ are then read off Table IV. The characteristic value is then determined from (87), (88), and (68), particular examples of which are given in (89) and (90) for the "exponential model," and in (94) and (96) for the "power law" model. These include a correction term Δ_m which was obtained from the perturbation theory of transitional modes developed in Section 4D. A comparison between exact characteristic values and those obtained from the theory of transitional modes is shown in Tables II and III for the "exponential" and "power law" models, respectively. It is seen that the theory of transitional modes developed here yields accurate characteristic values over a wide region, extending into the leaky modes on one side and into the trapped modes on the other side. The height-gain functions can be obtained from (60), (69), (76), (77), (83), and (101). In the case of an exponential model the variables u and v can be computed with the aid of Fig. 2. The asymptotic theory of trapped modes developed in Section 5 follows Langer's method, whereby the difficulty with the Stokes phenomenon is completely obviated. The procedure used is to develop the solutions around the two turning points h_1 and h_2 shown in Fig. 1 and to join the solutions at the duct height h_0 . The relation determining the characteristic values thus obtained is (140), which bears a similarity to the corresponding equation for a "bilinear" model. Indeed, Eqs. (131), (132), (136), and (137) give the transformation required to turn the model with a continuous $y(h)$ curve shown in Fig. 1 into one where $y(h)$ is represented by two straight lines that intersect at h_0 . It is shown in Eq. (159) how one can derive the Furry-Gamow formula for the decrement of *completely trapped modes* from Eq. (140). Equations (160) to (163) give expressions for the height-gain functions for highly trapped modes. These cover the whole range of height including the vicinity of the two turning points.

1. INTRODUCTION

THIS paper gives an account of several developments in the theory of asymptotic solutions for the normal modes in microwave propagation. In the theory of microwave propagation we are interested in determining the electromagnetic field generated by a dipole antenna which is situated at some elevation above

the ground. Of particular interest is the effect of atmospheric refraction on propagation into regions below the horizon. The field can be determined from a Hertzian potential Ψ^1 which satisfies the wave equation

$$\nabla^2 \Psi + k^2 \mu^2 \Psi = 0. \quad (1)$$

Here

$$k = 2\pi/\lambda, \quad (2)$$

$$k^2 \mu^2 = (\epsilon\omega^2 - i\sigma\omega)/c^2, \quad (3)$$

* This paper is based on work done for the Naval Research Laboratory under contract No. N5-ori-90 with the Navy's Office of Research and Inventions.

** Columbia University Mathematical Physics Group, on leave of absence from M.I.T.

¹ See Arnold Sommerfeld's article in Frank-Mises *Differentialgleichungen der Physik* (M. S. Rosenberg, New York), Vol. II, p. 918.

μ denotes the index of refraction and a time factor $e^{i\omega t}$ has been assumed. For vertical polarization Ψ represents the vertical component of the electric Hertzian potential, while for horizontal polarization it represents the vertical component of the magnetic Hertzian potential. The observed field intensity is proportional to $|\Psi|$ in each case. Equation (1) is to be solved subject to the following boundary conditions:

- (a) Ψ should reduce to the form e^{-ikR}/R for small distances R from the source,
- (b) Ψ should vanish at the ground, and
- (c) Ψ should represent an outgoing wave at great distances from the source.

The boundary condition (b) is not strictly true, but it is a good approximation for the low order normal-mode solutions of Eq. (1).

The quantity μ appearing in Eq. (1) represents the index of refraction of the air, which will be assumed to vary only with height above the ground, and the boundary condition (b) is to be satisfied at the spherical surface of the earth. The analysis can be simplified by the use of a device attributed originally to Schelleng, Burrows, and Ferrell² and later developed by M. H. L. Pryce,³ whereby space is transformed so as to make the surface of the earth plane and the rays curved. The transformation can be accomplished by using in Eq. (1) instead of μ the so-called *modified index of refraction* N defined by

$$N = r\mu(r)/b\mu(b),$$

where r denotes the distance from the center of the earth and b is the radius at some reference level. Taking $b = a$, the radius of the earth, we have

$$\frac{\mu(r)}{\mu(a)} = 1 + \frac{h}{a}, \quad (4)$$

$$N \cong 1 + h \left(\mu + \frac{1}{a} \right) + \mu'(h). \quad (5)$$

Here h denotes height above ground, μ the vertical gradient of the index of refraction in the *standard atmosphere*, and $\mu'(h)$ denotes the *refraction anomaly* or the deviation of the refractive index from standard conditions. The term $h\mu$ in Eq. (5) has the effect of modifying the radius of the earth a to an effective value a_e determined by

$$1/a_e = 1/a + \mu;$$

the value of a_e/a is generally taken as 4/3. In practice it is convenient to use the quantity M defined by

$$M - M_0 = 10^6(N - 1).$$

Substituting in Eq. (5) and using the dependence of μ on atmospheric pressure p (millibars), temperature T (absolute scale) and the partial pressure of water vapor e (millibars), one finds that

$$M = \frac{79p}{T} + \frac{3.8 \times 10^6 e}{T^2} + .048h,$$

where h is expressed in feet. Here the term $.048h$ is $10^6(h/a)$; under standard conditions the sum of the first two terms in the expression for M varies with height like $-.012h$, giving

$$M_{std} = M_0 + .036h.$$

The increase of M_{std} with height results in a curving of the rays away from the ground, in the framework of the flattened earth, the radius of curvature of the rays being constant in a standard atmosphere and equal to a_e . Under certain atmospheric conditions the modified index of refraction increases with elevation in the ground layer at a more rapid rate than under standard conditions. We then have a case of *substandard refraction*, and transmission is generally poorer than in the standard atmosphere. On the other hand, when the modified index decreases with elevation in the ground layer at a less rapid rate than in the standard atmosphere, or when the initial slope of $M(h)$ -curve is negative, we have a state of *super-refraction*, and transmission is generally improved. A state of super-refraction which starts at the ground is called a *surface-duct*; super-refraction may also occur at some elevation, in which case it is referred to as an *elevated duct*.

² J. C. Schelleng, C. R. Burrows, and E. B. Ferrell, Proc. I.R.E. 21, 427 (1933); J. E. Freehafer, Radiation Laboratory Report 447 (1943); C. L. Pekeris, Phys. Rev. 70, 518 (October, 1946).

³ Unpublished report.

already been fully dealt with by Furry⁴ we shall confine our discussion of these cases to certain extensions which have been developed subsequent to Furry's report. Thus, in order to have an estimate of the order of magnitude of the error involved in the use of an asymptotic solution for a given value of k , we derive explicit expressions for the terms after the leading one in the asymptotic expansion (in inverse fractional powers of k) of the solutions of (9). We also use the Langer⁵ method throughout and obtain solutions which do not suffer from the Stokes phenomenon, and which, furthermore, can be used to compute $U_m(h)$ in the vicinity of the turning points. The latter is, of course, required for determining the field. In addition, a direct method is given for computing the characteristic value from the "phase integral relation" which is particularly useful in case of substandard refraction. The principal contribution of this paper is however in the extension of the method of asymptotic solutions to cover *transitional modes*.

2. ASYMPTOTIC SOLUTIONS FOR LEAKY MODES

In case of leaky modes the imaginary part of the characteristic value Λ_m is comparable in magnitude to the real part, and the expression in brackets in the equation

$$d^2 U_m / dh^2 + k^2 [\Lambda_m + y(h)] U_m = 0 \quad (9')$$

does not vanish on the real axis of h . It was pointed out by Eckersley⁶ that asymptotic solutions of (9) can be obtained by determining a *turning point* h_1 (complex) defined by

$$\Lambda_m + y(h_1) = 0, \quad (11)$$

and developing the solutions around that point. Following Langer's method⁵ we write

$$U_m(h) = \frac{G(h)}{[Q(h)]^{1/2}}, \quad Q^2 = k^2 [\Lambda_m + y(h)], \quad (12)$$

$$u(h) = \int_{h_1}^h Q(h) dh$$

⁴ R. E. Langer, Phys. Rev. 51, 669 (1937); see also H. Jeffreys, Phil. Mag. 33, 451 (1943).

⁵ T. L. Eckersley, Proc. Roy. Soc. A132, 83 (1931); 136, 499 (1932); 137, 158 (1932); T. L. Eckersley and G. Millington, Phil. Trans. Roy. Soc. A237, 273 (1938).

and obtain from (9)

$$\frac{d^2 G}{dh^2} - \frac{\dot{Q}}{Q} \frac{dG}{dh} + \left(-\frac{d^2 Q / dh^2}{2Q} + \frac{3}{4} \frac{\dot{Q}^2}{Q^2} + Q^2 \right) G = 0, \quad (13)$$

where dots denote differentiation with respect to h . We now introduce u as an independent variable in place of h , and obtain from (13):

$$d^2 G(u) / du^2 + P(u) G(u) = 0, \quad (14)$$

$$P(u) = 1 + \frac{1}{4Q^2} \left(\frac{dQ}{du} \right)^2 - \frac{1}{2Q} \frac{d^2 Q}{du^2}. \quad (15)$$

From its definition in (12) we develop $u(h)$ into a power series in $(h - h_1)$, invert this series and then express Q as a power series in u . From the latter we find that

$$P(u) = 1 + \frac{5}{36u^2} + \frac{A}{k^{4/3} u^{2/3}} + \frac{B}{k^2} + \frac{Cu^{2/3}}{k^{8/3}} + \dots, \quad (16)$$

where

$$A = \frac{(1.5)^{4/3}}{315(\dot{y})^{2/3}} \left[9 \left(\frac{y}{\dot{y}} \right)^2 - 10 \left(\frac{d^2 y / dh^2}{\dot{y}} \right) \right], \quad (17)$$

$$B = \frac{1}{450\dot{y}} \left[-25 \left(\frac{d^4 y / dh^4}{y} \right) + 70 \left(\frac{y(d^3 y / dh^3)}{\dot{y}^2} \right) - 42 \left(\frac{\dot{y}}{y} \right)^3 \right], \quad (18)$$

and

$$\dot{y} = \frac{dy(h_1)}{dh}, \quad y = \frac{d^2 y(h_1)}{dh^2}, \quad \text{etc.} \quad (19)$$

Using (16) in (14) and expanding $G(u)$ in a descending power series in k we obtain

$$G(u) = u^{1/3} I_{1/3}^{(2)}(u) + \frac{3A}{2k^{4/3}} u^{5/3} I_{-2/3}^{(2)}(u) - \frac{B}{2k^2} u^{3/3} I_{4/3}^{(2)}(u) + O(k^{-5/3}) + \dots, \quad (20)$$

where we have chosen the Hankel function of the *second kind* in order to assure an outgoing wave at great heights.

The asymptotic expansion (20) of the solution of Eq. (14) is useful for moderate values of u and in the limit of large values of k . This combi-

nation of values of u and k actually occurs in many applications of (20), such as, for example, the finding of the zeros of $G(u)$, and the problem under consideration. Taking the boundary condition at the ground as

$$G[u(0)] = 0, \quad u(0) = k \int_{h_1}^0 [\Lambda_m + y(h)]^{1/2} dh, \quad (7')$$

we seek an expansion for $u(0)$ of the form

$$u(0) = c + \frac{c_1}{k^{1/3}} + \frac{c_2}{k^2} + \frac{c_3}{k^{5/3}} + \dots \quad (21)$$

Substituting in (20) we find

$$H_{1/3}^{(2)}(c) = 0; \quad c = e^{i\pi} v_m, \quad (22)$$

where the v_m are the roots of the equation

$$J_{1/3}(v_m) + J_{-1/3}(v_m) = 0, \quad v_m \simeq \pi(m - \frac{1}{4}), \\ m = 1, 2, 3, \dots \quad (22')$$

When this value of $u(0)$ is used in (7') we obtain *Eckersley's phase integral condition*:

$$k \int_0^{h_1} [\Lambda_m + y(h)]^{1/2} dh \\ = k \int_0^{h_1} [y(h) - y(h_1)]^{1/2} dh = v_m. \quad (23)$$

This gives a relation for determining the value (complex) of the turning point h_1 , from which Λ_m is then evaluated by Eq. (11). The numerical solution of (23) for h_1 is, however, a difficult problem, and no method of solution of general applicability is at present available, especially for the case of leaky modes. In the following section we develop an approximate, but direct, method of solving Eq. (23), which is effective in the case when $y(h)$ is a monotone function of h .

Proceeding now with the asymptotic expansion (21) we find that

$$c_1 = -\frac{3A}{2} \left[\frac{c^{1/3} H_{-2/3}^{(2)}(c)}{H_{1/3}^{(2)}(c)} \right] = -\frac{3A}{2} c^{1/3}, \quad (24)$$

$$c_2 = \frac{B}{2} \left[\frac{c H_{4/3}^{(2)}(c)}{H_{1/3}^{(2)}(c)} \right] = -\frac{Bc}{2}. \quad (25)$$

Hence the roots $u(0)$ of $G(u)$ in (20) are given by

$$u(0) = e^{i\pi} v_m - \frac{3A}{2k^{4/3}} e^{i\pi/2} v_m^{1/3} + \frac{Bv_m}{2k^2} + \dots \quad (26)$$

With this expansion, the phase integral condition (23) is replaced by

$$k \int_0^{h_1} [\Lambda_m + y(h)]^{1/2} dh \\ = v_m + \frac{3A}{2k^{4/3}} e^{i\pi/2} v_m^{1/3} - \frac{Bv_m}{2k^2} + \dots \quad (27)$$

Let now

$$\Lambda_m = \Lambda_m^{(0)} + \frac{\Delta}{k^{2/3}} + \dots, \quad (28)$$

substitute in (27) and obtain

$$k \int_0^{h_1} [\Lambda_m^{(0)} + y(h)]^{1/2} dh = v_m, \quad (23')$$

$$\Delta = \frac{3A}{\theta} e^{i\pi/3} (v_m)^{1/3}, \quad (29)$$

where

$$\theta = \int_0^{h_1} \frac{dh}{[\Lambda_m^{(0)} + y(h)]^{1/2}}. \quad (30)$$

To summarize, the determination of the characteristic values Λ_m in case of leaky modes proceeds as follows. First one finds the turning point h_1 and Λ_m by solving (23'), where

$$\Lambda_m^{(0)} = -y(h_1). \quad (11')$$

One then evaluates the derivatives of $y(h)$ at h_1 and finds A by (17), as well as θ from (30). Λ_m is then given by

$$\Lambda_m = \Lambda_m^{(0)} + \frac{3A}{\theta k^{2/3}} e^{i\pi/3} (v_m)^{1/3} + \dots \quad (28')$$

The last term indicates the order of magnitude of the error involved in the use of $\Lambda_m^{(0)}$ for Λ_m . It will be noted that the factor A contains inverse powers of $y(h)$, so that the accuracy of the phase integral method for determining the characteristic values depends on (a) k being large and on (b) the turning point h_1 being removed from a place where $y(h_1) = 0$. The last condition is violated, in case of a surface duct, when the mode becomes transitional, i.e., when h_1 approaches the real value of the "duct height" h_0 shown in Fig. 1.

As for the height-gain functions, we have in

the first approximation

$$U_m(h) = \frac{C_m u^{1/2} H_{1/2}^{(2)}(u)}{[y(h) + \Lambda_m]^{1/4}}, \quad (31)$$

where

$$u = k \int_{h_1}^h [y(h) + \Lambda_m]^{1/4} dh,$$

and the normalization factor C_m is to be determined from Eq. (10). We find that at $h=0$

$$\frac{\partial U_m}{\partial h} = C_m k [y(0) + \Lambda_m]^{1/4} u_0^{1/2} \dot{H}_{1/2}^{(2)}(u_0), \quad (32)$$

$$\frac{\partial U_m}{\partial \Lambda_m} = -\frac{1}{2} C_m k \theta [y(0) + \Lambda_m]^{-1/4} u_0^{1/2} \dot{H}_{1/2}^{(2)}(u_0), \quad (33)$$

$$\dot{H}_{1/2}^{(2)}(u_0) = \frac{2}{3^{1/2}} e^{i\pi/6} \left(\frac{2u_m}{v_m} \right)^{1/2};$$

$$u_m \equiv \frac{v_m}{2} [J_{2/3}(v_m) - J_{-2/3}(v_m)]^2 \frac{3}{\pi}. \quad (34)$$

Substituting in (10) we determine C_m , and then obtain

$$U(h) = \frac{e^{-i2\pi/3} (3u)^{1/2} H_{1/2}^{(2)}(u)}{2(\theta u_m)^{1/2} [y(h) + \Lambda_m]^{1/4}}. \quad (35)$$

One finds from (35) that

$$\frac{dU_m(0)}{dh} = k \left(\frac{2}{\theta} \right)^{1/2} [y(0) + \Lambda_m]^{1/4}. \quad (36)$$

In evaluating the height-gain function it is convenient to use the Harvard tables of the function

$$h_2(\chi) \equiv \left(\frac{2}{3} \right)^{1/2} \chi^{1/2} H_{1/2}^{(2)} \left(\frac{2}{3} \chi^{3/2} \right) = u^{1/2} H_{1/2}^{(2)}(u); \quad (37)$$

$$\chi = \left(\frac{3u}{2} \right)^{2/3},$$

which is tabulated for complex values of the argument χ . In terms of this function we have

$$U_m(h) = \frac{3^{1/2} u^{1/6} e^{-i2\pi/3} h_2(\chi)}{2(\theta u_m)^{1/2} [y(h) + \Lambda_m]^{1/4}}. \quad (38)$$

If we include the term after the leading one in (20), the function $H_{1/2}^{(2)}(u)$ in (35) is to be replaced by $[H_{1/2}^{(2)}(u) + (3A/2k^{4/3})u^{1/2}H_{-2/3}^{(2)}(u)]$, or the function $h_2(\chi)$ in (38) by $[h_2(\chi) + (3/2)^{2/3} \times (A/k^{4/3})h_2'(\chi)]$.

It may be pointed out that if we use the first three terms of $P(u)$ in (16) the *exact* solution of (14) is

$$G = u^{1/6} (u^{2/3} + A/k^{4/3})^{1/2} H_{1/2}^{(2)} [(u^{2/3} + A/k^{4/3})^{3/2}]$$

$$= u^{1/6} h_2 \left[\left(\frac{3u}{2} \right)^{2/3} + \left(\frac{3A^{2/3}}{2k^2} \right)^{2/3} \right]. \quad (39)$$

Equation (26) is then replaced by

$$[u(0)^{2/3} + A/k^{4/3}]^{3/2} = e^{i\pi} v_m, \quad (40)$$

and Eq. (29) by

$$\Delta = \frac{2v_m k^{4/3}}{\theta} [(1 + A e^{i\pi/3} / k^{4/3} v_m^{2/3})^{3/2} - 1]. \quad (41)$$

3. AN APPROXIMATE METHOD OF DETERMINING THE CHARACTERISTIC VALUES FROM THE PHASE INTEGRAL EQUATION

The problem of solving for the characteristic values Λ_m by the phase integral equation is to determine a turning point h_1 (generally complex) which satisfies the equation

$$F(h_1) = k \int_0^{h_1} [y(h) - y(h_1)]^{1/2} dh = v_m, \quad (42)$$

where $y(h)$ is a given continuous function representing the variation of the modified refractive index with height, and v_m is a known constant. The method of solving (42) to be described presently is particularly suited to the case when the function $y(h)$ is monotonic, so that h is a single-valued function of y . The idea is to write

$$\int_0^{h_1} [y(h) - y(h_1)]^{1/2} dh$$

$$= \int_{y(0)}^{y(h_1)} [y - y(h_1)]^{1/2} \left(\frac{dh}{dy} \right) dy, \quad (43)$$

and consider (dh/dy) as a function of y . Let

$$[y(0) + \Lambda_m]^{1/2} = [y(0) - y(h_1)]^{1/2} \equiv x; \quad (44)$$

then by integrating partially with respect to y we obtain from (42)

$$-\frac{2}{3} h x^2 + \frac{4}{15} \left(\frac{d^2 h}{dy^2} \right) x^5 - \frac{8}{105} \left(\frac{d^3 h}{dy^3} \right) x^7$$

$$+ \frac{16}{945} \left(\frac{d^4 h}{dy^4} \right) x^9 - \dots = v_m/k. \quad (45)$$

Equation (45) is now inverted to express x as a power series in $e^{i\pi/3}(v_m)^{1/3}$. Upon squaring the latter series we obtain, on using (44),

$$\Lambda_m = x^2 - y(0) = -y(0) + \tau^2 + \frac{4}{15}h_2\tau^4 + \tau^6 \left[\frac{4}{25}h_2^2 - \frac{8}{105}h_1 \right] + \frac{16}{27}\tau^8 \left[\frac{77}{375}h_2^3 - \frac{33}{175}h_2h_3 + \frac{1}{35}h_4 \right] + \dots, \quad (46)$$

where

$$h_2 = \left(\frac{d^2h/dy^2}{h} \right), \quad h_3 = \left(\frac{d^3h/dy^3}{h} \right), \quad \text{etc.}, \quad (47)$$

$$\tau = e^{i\pi/3}(3v_m/2kh)^{1/3}. \quad (48)$$

When h is positive the phase of τ is $\pi/3$; in case of a surface duct the phase of τ is zero. In the latter case Eq. (46) yields a real value for Λ_m .

Let $y(h)$ be given as a power series in h :

$$y(h) = y(0) + b_1h + b_2h^2 + \dots + b_nh^n + \dots, \quad (49)$$

then

$$h = \frac{dh}{dy}(h=0) = 1/b_1, \quad \frac{d^2h}{dy^2} = -2b_2/b_1^3, \quad (50)$$

$$d^3h/dy^3 = (-6b_1b_3 + 12b_2^2)/b_1^5, \quad (51)$$

$$d^4h/dy^4 = (120b_1b_2b_3 - 24b_1^2b_4 - 120b_2^3)/b_1^7, \quad \text{etc.} \quad (52)$$

Equation (46) then gives an explicit expression for Λ_m in terms of the coefficients in the power series expansion (49) of the modified index of refraction curve $y(h)$. This expression is of the nature of an asymptotic formula; it should be terminated when the individual terms begin to increase, and the error in Λ_m is then of the order of magnitude of the last term retained.

In an actual case the value of Λ_m obtained from (46) should be checked by computing h_1 from $y(h_1) = -\Lambda_m$, or from

$$h_1 = -hx + \frac{d^2h/dy^2}{2!}x^2 - \frac{d^3h/dy^3}{3!}x^3 + \dots, \quad (53)$$

and evaluating the phase integral $F(h_1)$ in (42). An improved value of h_1 can then be obtained

TABLE I. Approximate determination of Λ_1 from (56) and the verification that $\int_0^{h_1} [y(h) + \Lambda_1]^{1/2} dh = 2.383$.

α	λ	Λ_1 from (56)	h_1 from $y(h_1) = -\Lambda_1$	$\int_0^{h_1} [y(h) + \Lambda_1]^{1/2} dh$	v_1
-20	.6156	12 360 + 9 775i	.3997 - .9518i	2 370 + 004i	2 383
-10	.6356	4 878 + 6 302i	.4745 - 1.1982i	2 397 + 010i	2 383
-5	.6356	1 499 + 4 257i	.5691 - 1.4692i	2 393 + 009i	2 383
-2	.6356	-246 + 2 902i	-.764 - 1.787i	2 363 - 008i	2 383

by Newton's method, using the relation

$$\frac{dF(h_1)}{dh_1} = -\frac{k}{2y(h_1)} \int_0^{h_1} \frac{dh}{[y(h) - y(h_1)]^{1/2}} \quad (54)$$

The degree of accuracy that can be obtained by the method described above will now be illustrated for the case of an exponential model, where

$$y(h) = h + \alpha e^{-\lambda h}, \quad k = 1. \quad (55)$$

In this case Eq. (46) reduces to

$$\Lambda_m = -\alpha + \tau^2 - \frac{4}{15} \frac{\alpha\lambda^2}{(1-\alpha\lambda)^2} \tau^4 + \frac{4}{525} \frac{\alpha\lambda^3(\alpha\lambda - 10)}{(1-\alpha\lambda)^4} \tau^6$$

$$- \frac{16\alpha\lambda^4\tau^8}{70875(1-\alpha\lambda)^6} (75 + 105\alpha\lambda - \alpha^2\lambda^2) + \dots \quad (56)$$

$$\tau = e^{i\pi/3} [3v_m(1-\alpha\lambda)/2]^{1/3}. \quad (57)$$

It is seen from Table I that the values for Λ_1 obtained from (56) for the cases treated satisfy the phase integral equation rather closely. Another case treated was $\alpha = +20$, $\lambda = .6356$, where a real value for Λ_1 of -10.22 was obtained from (56). This is to be compared with the exact value of $\Lambda_1 = -10.21 + 1.07 \times 10^{-14}i$ obtained for this case by Pearcey and Whitehead.

4. ASYMPTOTIC SOLUTIONS FOR TRANSITIONAL MODES

A. Formal Solution

In the case of transitional modes the characteristic value Λ_m has a small imaginary part, the real part being very nearly equal to $-y(h_0)$, where h_0 denotes the duct height (see Fig. 1). The quantity in brackets in

$$\frac{d^2U_m}{du^2} + k^2[\Lambda_m + y(h)]U_m = 0, \quad (9')$$

* The specification of $k=1$ means simply that we are using natural units, which one would do in an actual problem.

therefore vanishes in the complex h plane at a point h_1 (turning point) which is very close to h_0 on the real axis. Since h_0 is defined by

$$\dot{y}(h_0) = 0, \quad (58)$$

the proximity of h_1 to h_0 implies that the asymptotic expansion (20), which proceeds in inverse powers of $\dot{y}(h_1)$, breaks down for transitional modes.

Our method of dealing with the case of transitional modes is to develop the solution around y_0 as a turning point.† In place of (12) we now write

$$Q^2 = k^2[y(h) + \Lambda_m^0], \quad \Lambda_m^0 = -y(h_0), \quad (59)$$

$$u(h) = k \int_{h_0}^h [y(h) - y(h_0)]^{\frac{1}{2}} dh, \quad (60)$$

$$U_m(h) = \frac{G(h)}{[Q(h)]^{\frac{1}{2}}},$$

and find that

$$\frac{d^2 U_m}{dh^2} + [Q^2(h) + k^2(\Lambda_m - \Lambda_m^0)] U_m = 0, \quad (61)$$

$$\frac{d^2 G}{dh^2} - \frac{\dot{Q}}{Q} \frac{dG}{dh} + \left[-\frac{d^2 Q/dh^2}{2Q} + \frac{3}{4} \frac{\dot{Q}^2}{Q^2} + Q^2 + k^2(\Lambda_m - \Lambda_m^0) \right] G = 0, \quad (62)$$

$$\frac{d^2 G(u)}{du^2} + P(u)G(u) = 0, \quad (63)$$

$$P(u) = 1 + \frac{1}{4Q^2} \left(\frac{dQ}{du} \right)^2 - \frac{1}{2Q} \frac{d^2 Q}{du^2} + \frac{k^2(\Lambda_m - \Lambda_m^0)}{Q^2}. \quad (64)$$

As in Eq. (13), the dots in (62) denote differentiation with respect to h .

In contrast to the case of leaky modes the quantities u , Q , and Λ_m^0 are now real.

† Subsequent to the completion of this investigation the writer learned of a paper by S. Goldstein, Proc. Lond. Math. Soc. 33, 246 (1932), where the same idea had been used in another connection. Also, Dr. T. Pearcey informed the writer that he independently treated the transitional case by a similar method.

Next, we proceed with the development of $P(u)$ into a power series in u . This is accomplished by first expressing u as a power series in $(h - h_0)$ using (60), inverting this series, and then using it in (59) to obtain a power series in u for Q . When substituted in (64) it yields

$$P(u) = 1 + \frac{3}{16u^2} + \frac{\gamma}{u} + \frac{N}{u^{\frac{1}{2}}} + \dots, \quad (65)$$

$$\gamma = k(\Lambda_m - \Lambda_m^0)/[2y(h_0)]^{\frac{1}{2}} + \frac{1}{576c}(7y_3^2 - 9y_4), \quad (66)$$

$$N = -\frac{4}{9} \frac{y_3 \gamma}{(2c)^{\frac{1}{2}}} + \frac{1}{97200 \sqrt{2} c^{\frac{3}{2}}} \times (-1944y_6 + 3645y_3 y_4 - 1715y_3^3), \quad (67)$$

where

$$c = k[y_0(h)/2]^{\frac{1}{2}}, \quad y_3 = \left[\frac{d^2 y(h_0)}{dy^2} \right] / y(h_0), \quad (68)$$

$$y_4 = \left[\frac{d^4 y(h_0)}{dy^4} \right] / y(h_0), \text{ etc.}$$

This expansion for $P(u)$ is to be used for $h > h_0$, with u defined in (60). For $0 < h < h_0$ it is convenient to introduce the variable

$$v = k \int_h^{h_0} [y(h) - y(h_0)]^{\frac{1}{2}} dh, \quad (69)$$

and to develop P into a power series in v . In doing that, one must keep in mind that Q is a positive function of v , as it is of u . One thus finds that

$$P(v) = 1 + \frac{3}{16v^2} + \frac{\gamma}{v} - \frac{N}{v^{\frac{1}{2}}} + \dots^{**} \quad (70)$$

It will be shown later that the product $k(\Lambda_m - \Lambda_m^0)$, and with it γ , remains finite in the limit of large k , so that N is of the order of $1/k^{\frac{1}{2}}$ relative to γ . We therefore treat N as a small perturbation term and write

$$G(u) = W(u) + Nw(u) + \dots, \quad (71)$$

A comparison of (70) with (65) shows that in going from $h > h_0$ to $h < h_0$ the phase of u can be considered as undergoing a change of 2π .

where

$$d^2W(u) + \left(1 + \frac{3}{16u^2} + \frac{\gamma}{u}\right)W(u) = 0, \quad h > h_0, \quad (72)$$

$$\frac{d^2W(v)}{dv^2} + \left(1 + \frac{3}{16v^2} + \frac{\gamma}{v}\right)W(v) = 0, \quad h < h_0, \quad (73)$$

$$\frac{d^2w(u)}{du^2} + \left(1 + \frac{3}{16u^2} + \frac{\gamma}{u}\right)w(u) = -\frac{W(u)}{u^{\frac{1}{2}}}, \quad h > h_0, \quad (74)$$

$$\frac{d^2w(v)}{dv^2} + \left(1 + \frac{3}{16v^2} + \frac{\gamma}{v}\right)w(v) = +\frac{W(v)}{v^{\frac{1}{2}}}, \quad h < h_0. \quad (75)$$

The solutions of (72) and (73) are confluent hypergeometric functions of arguments $2iu$ and $2iv$, respectively. Since the phase of $W(u)$ must decrease for large u , the appropriate solution of (72) is

$$W_m(u) = C_m W_{\kappa, 1/4}(2iu), \quad \kappa = -i\gamma/2, \quad (76)$$

where $W_{\kappa, m}(z)$ is Whittaker's function⁷

$$W_{\kappa, 1/4}(2iu) = \frac{2\pi^{\frac{1}{2}}}{\Gamma(\frac{1}{4} - \kappa)} M_{\kappa, 1/4}(2iu) + \frac{\pi^{\frac{1}{2}}}{\Gamma(\frac{1}{4} - \kappa)} M_{\kappa, -1/4}(2iu), \quad (77)$$

$$M_{\kappa, 1/4}(z) = z^{1/4} e^{-z/2} F\left(\frac{3}{4} - \kappa, \frac{3}{2}; z\right); \quad (78)$$

$$M_{\kappa, -1/4}(z) = z^{1/4} e^{-z/2} F\left(\frac{1}{4} - \kappa, \frac{1}{2}; z\right).$$

For large z , $W_{\kappa, m}(z)$ behaves like

$$W_{\kappa, m}(z) \rightarrow z^{\frac{1}{2}} e^{-z/2} F\left(\frac{3}{4} - \kappa, \frac{1}{4} - \kappa; -\frac{1}{z}\right), \quad -\pi < \arg z < \pi, \quad (79)$$

$$W_m(u) \rightarrow C_m \{1 + [\frac{1}{16} - (\kappa - \frac{1}{2})^2]/2iu + \dots\} \times \exp\{-iu + \kappa[\ln(2u) + i\pi/2]\}.$$

In the region below the duct height ($h < h_0$) we set

$$W_m(v) = C_m [p M_{\kappa, 1/4}(2iv) + q M_{\kappa, -1/4}(2iv)], \quad (80)$$

determine the constants p and q from the

⁷E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, New York, 1935), p. 339.

conditions that at $h = h_0$ ($u = v = 0$)

$$U_m(u) = \frac{W_m(u)}{[Q(u)]^{\frac{1}{2}}} = \frac{W_m(v)}{[Q(v)]^{\frac{1}{2}}} = U_m(v), \quad (81)$$

and

$$\frac{d}{dh} U_m(u) = Q(u) \frac{d}{du} U_m(u) = \frac{d}{dh} U_m(v) = -Q(v) \frac{d}{dv} U_m(v). \quad (82)$$

We shall omit the details of the calculation here, and shall merely state that from condition (81) it follows that $q = \pi^{\frac{1}{2}}/\Gamma(\frac{3}{4} - \kappa)$, while condition (82) requires that $p = 2\pi^{\frac{1}{2}}/\Gamma(\frac{1}{4} - \kappa)$. Hence we have

$$W_m(v) = C_m \left[\frac{2\pi^{\frac{1}{2}}}{\Gamma(\frac{1}{4} - \kappa)} M_{\kappa, 1/4}(2iv) + \frac{\pi^{\frac{1}{2}}}{\Gamma(\frac{3}{4} - \kappa)} M_{\kappa, -1/4}(2iv) \right]. \quad (83)$$

B. Determination of the Characteristic Values

The characteristic values Λ_m are now determined by the boundary condition (8) at the ground.

$$P(\gamma, v_0) \equiv \frac{2\pi^{\frac{1}{2}}}{\Gamma(\frac{1}{4} - \kappa)} M_{\kappa, 1/4}(2iv_0) + \frac{\pi^{\frac{1}{2}}}{\Gamma(\frac{3}{4} - \kappa)} M_{\kappa, -1/4}(2iv_0) = 0, \quad (84)$$

$$v_0 = k \int_0^{h_0} [\gamma(h) - \gamma(h_0)]^{\frac{1}{2}} dh. \quad (85)$$

For a given value of v_0 , as determined from the modified refractive index curve $\gamma(h)$, Eq. (84) yields characteristic values (complex) of $\kappa_m = -i\gamma_m/2$, from which Λ_m is then evaluated by Eq. (66). The values of Λ_m thus obtained will be the more accurate the smaller the values of γ_m .

In order to have an estimate of the error to which the values γ_m^0 obtained from Eq. (84) are subject, we derive in section D below a first-order correction term arising from the $w(u)$ term in (71). The result is

$$\gamma_m = \gamma_m^0 - N(v_0) \frac{d\gamma_m^0}{dv_0}, \quad (86)$$

where N is defined in (67), and $d\gamma_m^0/dv_0$ is the slope of the $\gamma_m^0(v_0)$ curve resulting from (84).

Substituting this result in (66) we arrive at an expression for the characteristic values:

$$\Lambda_m = \Lambda_m^1 + \Delta_m;$$

$$\Lambda_m^1 = -y(h_0) + \frac{[2\bar{y}(h_0)]^{\frac{1}{2}}}{k} \gamma_m^0 - \frac{1}{576c} (7y_3^2 - 9y_4), \quad (87)$$

$$\Delta_m = -\frac{[2\bar{y}(h_0)]^{\frac{1}{2}}}{k} N(v_0) \frac{d\gamma_m^0}{dv_0}. \quad (88)$$

Here, γ_m^0 is to be used for the γ appearing in the expression for N . The term Δ serves not only as an indicator of the order of magnitude of the error in the leading term, but, as will be shown presently, considerably reduces the error when it is not large.

Table II shows how the characteristic values Λ_m computed from (87) compare with the exact values computed by Professor Hartree's associates on the differential analyzer for the case of the exponential model defined in Eq. (55). In this case Eqs. (87) and (88) reduce to

$$\Lambda_m = \Lambda_m^1 + \Delta_m;$$

$$\Lambda_m^1 = -(1 + \ln \alpha \lambda) / \lambda + (\lambda^2 / 144) + (2\lambda)^{\frac{1}{2}} \gamma_m^0, \quad (89)$$

$$\Delta_m = -(2\lambda v_0)^{\frac{1}{2}} \frac{d\gamma_m^0}{dv_0} - (.3737\lambda^{\frac{3}{4}} \gamma_m^0 + .000171\lambda^{\frac{9}{4}}). \quad (90)$$

The quantity v_0 defined in (85) can be obtained from

$$v_0 = f(x) / \lambda^{\frac{1}{2}}, \quad x = \ln \alpha \lambda, \quad (91)$$

$$f(x) = \int_0^x (e^z - 1 - z)^{\frac{1}{2}} dz.$$

The universal function $f(x)$ is plotted in Fig. 2. With the aid of this function one can also obtain u and v in (60) and (69) for the purpose of computing the height-gain functions from Eqs. (76) and (83).

A similar comparison is shown in Table III for the "Power law" model defined by

$$y(h) = h - 5h^{\frac{1}{2}}, \quad k = G^{\frac{1}{2}}. \quad (92)$$

Here

$$h_0 = 1, \quad v_0 = .4591G^{\frac{1}{2}}. \quad (93)$$

$$\Lambda_m = \Lambda_m^1 + \Delta_m; \quad (94)$$

$$\Lambda_m^1 = 4 + .5807\gamma_m^0/v_0 + .0166/v_0^2,$$

$$\Delta_m = -\frac{d\gamma_m^0}{dv_0} \left[\frac{.2798\gamma_m^0}{v_0} + \frac{.0370}{v_0^2} \right]. \quad (95)$$

TABLE II. Comparison of exact characteristic values Λ_m with values obtained from the theory of transitional modes in case of an exponential model. The exact values were computed by Professor Hartree's group on the differential analyzer.

α	λ	Mode No. m	v_0	Λ_m^1 (89)	$\Lambda_m^1 + \Delta_m$ (90)	Λ_m exact
10	1.2713	1	2.2374	-2.724 + .154i	-2.727 + .198i	-2.72 + .187i
10	.5538	1	3.0894	-5.038 + .019i	-5.115 + .018i	-5.123 + .016i
15	1.5840	1	2.7676	-2.946 + .066i	-3.080 + .069i	-3.10 + .06i
15	.8074	2	5.2285	-5.898 + .097i	-5.897 + .111i	-5.88 + .12i
15	.3024	2	5.8603	-8.423 + .029i	-8.440 + .029i	-8.44 + .026i
20	2.6458	1	2.3234	-1.880 + .193i	-1.880 + .281i	-1.88 + .27i
20	1.7886	1	3.1596	-3.244 + .027i	-3.610 + .014i	-3.74 + .008i
20	.9979	2	4.7089	-3.704 + .226i	-3.867 + .324i	-3.68 + .31i
20	.6357	2	6.3283	-5.955 + .018i	-6.066 + .016i	-6.07 + .01i
20	.6357	3	6.3283	-4.892 + .872i	-4.884 + .807i	-4.95 + .77i
20	.4307	3	7.6663	-7.071 + .171i	-7.060 + .230i	-7.00 + .2i
20	.2969	3	8.6816	-9.389 + .042i	-9.394 + .044i	-9.39 + .04i
20	.1928	3	8.6718	-12.364 + .019i	-12.274 + .020i	-12.27 + .02i

Tables II and III show that the method developed here for computing characteristic values is not only effective for strictly transitional modes but is also applicable over a considerable region extending into the leaky modes on the one side and into the trapped modes on the other side.

Values of γ_m^0 derived from Eq. (84) are shown in Table IV for the first three modes. In computing these, the Γ -functions were obtained from Stirling's asymptotic series⁸ after raising the arguments by a few units. The M -functions were computed from the power series (78) for values of v_0 up to 5.5. For larger values of v_0 the following asymptotic expansions were used

$$M_{\kappa, 1/4}(z) \rightarrow \frac{\pi^{\frac{1}{2}} z^{-\kappa} e^{-z/2}}{2\Gamma(\frac{3}{4} - \kappa)} F\left(\frac{3}{4} + \kappa, \frac{1}{4} + \kappa; -\frac{1}{z}\right) + \frac{\pi^{\frac{1}{2}} z^{\kappa} e^{-z/2}}{2\Gamma(\frac{3}{4} + \kappa)} \exp[-i\pi(\kappa - \frac{3}{4})] \times F\left(\frac{1}{4} - \kappa, \frac{3}{4} - \kappa; -\frac{1}{z}\right), \quad (96)$$

$$M_{\kappa, -1/4}(z) \rightarrow \frac{\pi^{\frac{1}{2}} z^{-\kappa} e^{-z/2}}{\Gamma(\frac{1}{4} - \kappa)} F\left(\frac{1}{4} + \kappa, \frac{3}{4} + \kappa; -\frac{1}{z}\right) + \frac{\pi^{\frac{1}{2}} z^{\kappa} e^{-z/2} \exp[-i\pi(\kappa - \frac{1}{4})]}{\Gamma(\frac{1}{4} + \kappa)} \times F\left(\frac{3}{4} - \kappa, \frac{1}{4} - \kappa; -\frac{1}{z}\right), \quad (97)$$

which are valid when $-\pi/2 < \arg z < 3\pi/2$. With

⁸ Reference 7, p. 251.

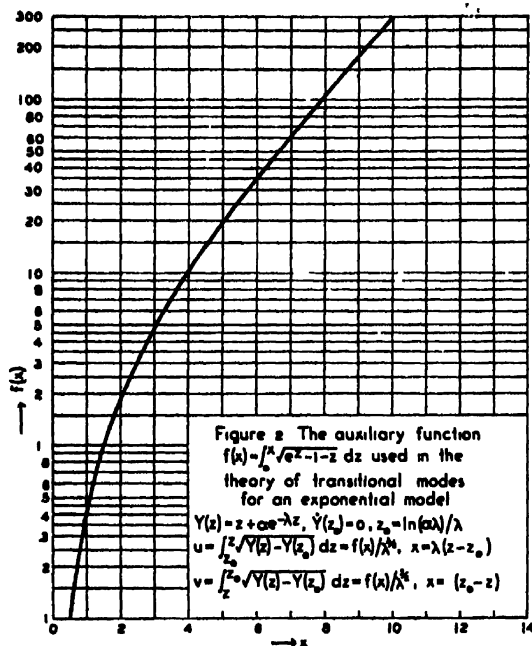


FIG. 2.

the aid of these asymptotic series one finds that for large v_0

$$W_m(v) \rightarrow C_m \{1 - \exp[-F_m(\gamma, v)]\} \times \exp \left[\kappa \ln(2v) + i \left(\frac{\pi}{2} - \frac{3\pi\kappa}{2} - v \right) - \frac{i\kappa(1-\kappa)}{2v} + \dots \right], \quad (98)$$

where

$$F_m(\gamma, v) = (1 + i\gamma) \ln \left(\frac{3}{2} + i\gamma \right) - \ln \left(\frac{1}{2} + i\gamma \right) - \gamma \left[\frac{\pi}{2} + i(1 + \ln 4 + \ln v) \right] + \left[-\frac{3}{2} + i \left(-2v - \frac{\pi}{2} + 2m\pi \right) \right] + \frac{1}{(18 + i12\gamma)} + \dots \quad (99)$$

Equation (84) is, therefore, replaced by $F_m(\gamma, v_0) = 0$.

C. Evaluation of the Normalization Factors

For the purpose of evaluating the height-gain functions we need to know the normalization constants C_m in (76) and (83). These constants are determined from the normalization condition

$$1 = -\frac{1}{k^2} \frac{\partial U_m(0)}{\partial h} \frac{\partial U_m(0)}{\partial \Lambda_m}. \quad (10')$$

With $U_m(h) = W_m(v)/[Q(v)]^{1/2}$, $\partial/\partial h = -Q(\partial/\partial v)$, we have

$$\frac{\partial U_m(0)}{\partial h} = -[Q(v_0)]^{1/2} \frac{\partial W_m(v_0)}{\partial v};$$

$$\frac{\partial U_m(0)}{\partial \Lambda_m} = \frac{k}{[2\ddot{y}(h_0)]^{1/2}} \frac{1}{[Q(v_0)]^{1/2}} \frac{\partial W_m(v_0)}{\partial \gamma_m}. \quad (100)$$

Substituting in (10') we get

$$C_m^2 = \frac{k[2\ddot{y}(h_0)]^{1/2}}{\left[\frac{\partial P(\gamma, v_0)}{\partial v} \frac{\partial P(\gamma, v_0)}{\partial \gamma} \right]}, \quad (101)$$

where $P(\gamma, v)$ is defined in (84). Values of $[(\partial P(v_0)/\partial v) \cdot (\partial P(v_0)/\partial \gamma)]$ are shown in Table IV. These were computed by the use of the asymptotic formulae (98) and (99), from which

TABLE III. Comparison of exact characteristic values Λ_m with values obtained from the theory of transitional modes in case of a power law model. The exact values were computed by Professor Hartree's group on the differential analyzer.

G^1	v_0	Mode No. m	Λ_m^1 (94)	$\Lambda_m^1 + \Delta_m$ (95)	Λ_m exact
10.242	4.701	1	3.841 + 0i	3.791 + 0i	3.749 + 0i
9.126	4.190	1	3.867 + 0i	3.828 + 0i	3.802 + 0i
8.134	3.734	1	3.894 + 0.005i	3.866 + 0i	3.851 + 0.0012i
7.249	3.328	1	3.921 + 0.017i	3.902 + 0.014i	3.879 + 0.0107i
6.458	2.964	1	3.948 + 0.043i	3.938 + 0.041i	3.942 + 0.0350i
5.758	2.643	1	3.974 + 0.102i	3.969 + 0.115i	3.978 + 0.146i
5.132	2.356	1	4.000 + 0.193i	4.001 + 0.238i	4.013 + 0.3304i
4.574	2.100	1	4.027 + 0.325i	4.031 + 0.415i	4.045 + 0.5507i
4.077	1.871	1	4.055 + 0.521i	4.060 + 0.68i	4.073 + 0.76i
3.633	1.668	1	4.084 + 0.744i	4.089 + 0.98i	4.098 + 1.08i
3.239	1.487	1	4.116 + 1.01i	4.119 + 1.35i	4.119 + 1.50i
2.886	1.325	1	4.147 + 1.40i	4.143 + 1.87i	4.140 + 1.98i
2.572	1.181	1	4.180 + 1.83i	4.166 + 2.44i	4.157 + 2.54i
2.292	1.052	1	4.222 + 2.32i	4.195 + 3.10i	4.169 + 3.16i
16.23	7.450	2	3.932 + 0i	3.914 + 0i	3.907 + 0i
14.46	6.640	2	3.958 + 0.008i	3.950 + 0.006i	3.949 + 0.004i
12.89	5.918	2	3.984 + 0.031i	3.981 + 0.032i	3.984 + 0.031i
11.49	5.275	2	4.007 + 0.099i	4.008 + 0.120i	4.012 + 0.151i
10.24	4.701	2	4.030 + 0.221i	4.032 + 0.280i	4.035 + 0.334i
9.126	4.190	2	4.052 + 0.406i	4.054 + 0.525i	4.053 + 0.600i
8.134	3.734	2	4.076 + 0.661i	4.075 + 0.862i	4.070 + 0.967i
7.249	3.328	2	4.102 + 0.991i	4.097 + 1.30i	4.088 + 1.43i
16.23	7.450	3	4.026 + 0.17i	4.027 + 0.22i	4.028 + 0.25i
14.46	6.640	3	4.047 + 0.36i	4.048 + 0.47i	4.044 + 0.50i

it follows that

$$\frac{\partial P(\gamma, v_0)}{dv} = -2i \left[1 + \frac{\gamma}{2v_0} - \frac{\gamma^2}{8v_0^2} \right] \exp \left[\kappa \ln(2v_0) \right] + i \left(\frac{\pi}{2} - \frac{3\pi\kappa}{2} - v_0 \right) - \frac{i\kappa(1-\kappa)}{2v_0}, \quad (102)$$

$$\frac{\partial P(\gamma, v_0)}{\partial \gamma} = \frac{\partial F_m(\gamma, v_0)}{\partial \gamma} \exp \left[\kappa \ln(2v_0) \right] + i \left(\frac{\pi}{2} - \frac{3\pi\kappa}{2} - v_0 \right) - \frac{i\kappa(1-\kappa)}{2v_0}, \quad (103)$$

where $F_m(\gamma, v)$ is defined in (99).

D. Perturbation Theory for Transitional Modes

In this section we shall derive the result stated in Eq. (86), namely that the correction to γ_m arising from the neglected $w(u)$ term in (71) is $-N(v_0) d\gamma_m^0/dv_0$ to within the first power in N . Actually it is possible to develop a perturbation theory in which N is not assumed to be small, as will be shown presently. Let Eq. (63) and its companion be written as

$$\frac{d^2 G}{du^2} + \left[1 + \frac{3}{16u^2} + \frac{\gamma}{u} + f(u) \right] G = 0, \quad h > h_0 \quad (104)$$

$$\frac{d^2 G}{dv^2} + \left[1 + \frac{3}{16v^2} + \frac{\gamma}{v} + g(v) \right] G = 0, \quad h < h_0 \quad (105)$$

and set

$$G(u) = \sum_{n=1}^{\infty} A_n W_n(u), \quad G(v) = \sum_{n=1}^{\infty} A_n W_n(v), \quad (106)$$

where

$$\frac{d^2 W_n(u)}{du^2} + \left[1 + \frac{3}{16u^2} + \frac{\gamma_n}{u} \right] W_n(u) = 0, \quad (72')$$

$$\frac{d^2 W_n(v)}{dv^2} + \left[1 + \frac{3}{16v^2} + \frac{\gamma_n}{v} \right] W_n(v) = 0. \quad (73')$$

When (106) is substituted in (104) and (105) we get

$$\sum A_n W_n(u) \left[\frac{\gamma - \gamma_n}{u} + f(u) \right] = 0, \quad (107)$$

$$\sum A_n W_n(v) \left[\frac{\gamma - \gamma_n}{v} + g(v) \right] = 0. \quad (108)$$

TABLE IV. Values of auxiliary functions used in the theory of transitional modes for the determination of characteristic values and normalization factors.

v_0 (85)	Mode No. m	γ_m^0 (84)	$\frac{d\gamma_m^0}{dv_0}$ (86)	$\left[\frac{\partial P(v_0)}{\partial v_0}, \frac{\partial P(v_0)}{\partial \gamma_m^0} \right]$ (102) and (103)
1.0	1	+ .386 + .439 <i>i</i>	- .160 - .458 <i>i</i>	+ 1.67 + 1.71 <i>i</i>
1.5	1	+ .276 + .255 <i>i</i>	- .265 - .302 <i>i</i>	2.23 + 3.34 <i>i</i>
2.0	1	+ .122 + .132 <i>i</i>	- .339 - .195 <i>i</i>	3.88 + 6.03 <i>i</i>
2.5	1	- .067 + .057 <i>i</i>	- .409 - .110 <i>i</i>	9.03 + 10.69 <i>i</i>
3.0	1	- .290 + .0191 <i>i</i>	- .482 - .048 <i>i</i>	26.1 + 18.2 <i>i</i>
3.5	1	- .551 + .0047 <i>i</i>	- .553 - .014 <i>i</i>	85.0 + 25.3 <i>i</i>
4.0	1	- .843 + .0008 <i>i</i>	- .610 - .003 <i>i</i>	306.7 + 4.94 <i>i</i>
4.5	1	- 1.161 + .0001 <i>i</i>	- .651 - .0004 <i>i</i>	1200. + 296 <i>i</i>
5.0	1	- 1.496 + .0 <i>i</i>	- .681 - .0 <i>i</i>	5146. + 2074 <i>i</i>
5.5	1	- 1.845 + .0 <i>i</i>	- .704 - .0 <i>i</i>	24250. + 11610 <i>i</i>
3.0	2	+ .651 + .695 <i>i</i>	- .261 - .398 <i>i</i>	+ .868 + 1.37 <i>i</i>
3.5	2	+ .539 + .500 <i>i</i>	- .229 - .356 <i>i</i>	.781 + 2.07 <i>i</i>
4.0	2	+ .420 + .340 <i>i</i>	- .244 - .284 <i>i</i>	.986 + 3.05 <i>i</i>
4.5	2	+ .292 + .216 <i>i</i>	- .269 - .215 <i>i</i>	1.52 + 4.66 <i>i</i>
5.0	2	+ .149 + .124 <i>i</i>	- .303 - .152 <i>i</i>	3.00 + 7.34 <i>i</i>
5.5	2	- .013 + .062 <i>i</i>	- .343 - .097 <i>i</i>	7.06 + 11.87 <i>i</i>
6.0	2	- .203 + .025 <i>i</i>	- .395 - .050 <i>i</i>	18.5 + 20.2 <i>i</i>
6.5	2	- .413 + .012 <i>i</i>	- .447 - .021 <i>i</i>	54.3 + 31.0 <i>i</i>
7.0	2	- .649 + .002 <i>i</i>	- .496 - .006 <i>i</i>	159.3 + 39.6 <i>i</i>
7.5	2	- .907 + .0005 <i>i</i>	- .535 - .001 <i>i</i>	503. + 3.9 <i>i</i>
8.0	2	- 1.183 + .0 <i>i</i>	- .567 - .0 <i>i</i>	1677. + 336 <i>i</i>
6.0	3	+ .695 + .613 <i>i</i>	- .244 - .330 <i>i</i>	+ .426 + 1.80 <i>i</i>
6.5	3	+ .566 + .453 <i>i</i>	- .248 - .291 <i>i</i>	.360 + 2.44 <i>i</i>
7.0	3	+ .442 + .319 <i>i</i>	- .251 - .243 <i>i</i>	.499 + 3.42 <i>i</i>
7.5	3	+ .314 + .211 <i>i</i>	- .261 - .192 <i>i</i>	.952 + 5.01 <i>i</i>
8.0	3	+ .179 + .131 <i>i</i>	- .283 - .141 <i>i</i>	2.27 + 7.59 <i>i</i>
8.5	3	+ .033 + .070 <i>i</i>	- .312 - .096 <i>i</i>	5.44 + 11.6 <i>i</i>
9.0	3	- .136 + .029 <i>i</i>	- .351 - .056 <i>i</i>	13.4 + 18.7 <i>i</i>
9.5	3	- .322 + .012 <i>i</i>	- .393 - .026 <i>i</i>	37.0 + 29.3 <i>i</i>
10.0	3	- .528 + .005 <i>i</i>	- .435 - .010 <i>i</i>	102. + 40.2 <i>i</i>
10.5	3	- .757 + .001 <i>i</i>	- .473 - .003 <i>i</i>	287. + 37.6 <i>i</i>
11.0	3	- 1.00 + .0 <i>i</i>	- .503 - .001 <i>i</i>	839. + 70 <i>i</i>

Upon multiplying (107) by $W_n(u)$ and integrating from 0 to ∞ , also upon multiplying (108) by $W_m(v)$ and integrating from 0 to v_0 , and then adding the results we get

$$\sum A_n \beta_{nm} = 0, \quad (109)$$

$$\beta_{nm} = (\gamma - \gamma_n) \left[\int_0^{\infty} W_n(u) W_m(u) \frac{du}{u} + \int_0^{v_0} W_n(v) W_m(v) \frac{dv}{v} \right] + \left[\int_0^{\infty} W_n(u) W_m(u) f(u) du + \int_0^{v_0} W_n(v) W_m(v) g(v) dv \right]. \quad (110)$$

Since the consistency condition for (109) requires that $|\beta_{nm}| = 0$, we have here a secular equation for the determination of γ , provided we can evaluate the elements β_{nm} . This can be accomplished for the case

$$f(u) = N/u^4, \quad g(v) = -N/v^4, \quad (111)$$

where it is found that

$$\beta_{nm} = (\gamma - \gamma_n) \dot{W}_n W_n' \delta_{nm} - N(v_0) \dot{W}_n \dot{W}_m / [1 + (\gamma_n - \gamma_m)^2], \quad (112)$$

$$\dot{W}_n \equiv \frac{\partial W_n(v_0)}{\partial v}, \quad W_n' \equiv \frac{\partial W_n(v_0)}{\partial \gamma}, \quad \dot{W}_n / W_n' = - \frac{d\gamma}{dv} \quad (113)$$

In a first-order perturbation theory the vanishing of β_{mm} yields Eq. (86).

The detailed derivation of (112) is rather lengthy, and we shall therefore limit ourselves here to an outline of the proof. To begin with it follows from (72') and (73') that

$$(\dot{W}_n W_m - W_n \dot{W}_m)_{u=0} = (\gamma_n - \gamma_m) \int_0^\infty W_n(u) W_m(u) \frac{du}{u}, \quad (114)$$

$$(\dot{W}_n W_m - W_n \dot{W}_m) \Big|_{v=1,0}^{v=0} = (\gamma_n - \gamma_m) \int_0^{v_0} W_n(v) W_m(v) \frac{dv}{v} \quad (115)$$

Also, it can be shown from (76), (77), and (83) that

$$[\dot{W}_n(v) W_m(v) - W_n(v) \dot{W}_m(v)]_{v=0} = 2i\pi C_n C_m \left[\frac{1}{\Gamma(\frac{1}{4} - \kappa_n) \Gamma(\frac{1}{4} - \kappa_m)} - \frac{1}{\Gamma(\frac{1}{4} - \kappa_m) \Gamma(\frac{1}{4} - \kappa_n)} \right] = -[\dot{W}_n(u) W_m(u) - W_n(u) \dot{W}_m(u)]_{u=0} \quad (116)$$

$$u \left(\frac{d^4 F}{du^4} \right) + \frac{d^3 F}{du^3} + \left[4u + \frac{3}{4u} + 2(\gamma_n + \gamma_m) \right] \left(\frac{d^2 F}{du^2} \right) + \left[4 - \frac{3}{2u^2} - \frac{(\gamma_n + \gamma_m)}{u} \right] \dot{F} + \left[\frac{3}{2u^3} + \frac{(\gamma_n + \gamma_m)}{u^2} + \frac{(\gamma_n - \gamma_m)^2}{u} \right] F = 0, \quad (122)$$

where we have written F for F_{nm} . This equation yields the useful relation

$$\gamma \frac{d}{du} \left\{ u^3 \left(u \frac{d^3 F}{du^3} - \frac{1}{2} \frac{d^2 F}{du^2} + \left[4u + \frac{1}{u} + 2(\gamma_n + \gamma_m) \right] \dot{F} + \left[-2 - \frac{1}{u^2} - \frac{2}{u} (\gamma_n + \gamma_m) \right] F \right) \right\} = -[1 + (\gamma_n - \gamma_m)^2] \frac{F}{u^3}, \quad (123)$$

Since $W_n(v_0) = W_m(v_0) = 0$, it follows that

$$\alpha_{nm} \equiv \int_0^{v_0} W_n(v) W_m(v) \frac{dv}{v} + \int_0^\infty W_n(u) W_m(u) \frac{du}{u} = 0, \quad n \neq m. \quad (117)$$

In order to obtain the value of α_{mm} we pass to the limit of $n \rightarrow m$ in the relation

$$(\gamma_n - \gamma_m) \alpha_{nm} = -(\dot{W}_n W_m - W_n \dot{W}_m)_{v=v_0}, \quad (118)$$

which follows from adding (114) and (115), and using (116). The result is

$$\alpha_{mm} = \int_0^{v_0} W_m^2(v) \frac{dv}{v} + \int_0^\infty W_m^2(u) \frac{du}{u} = \dot{W}_m \cdot W_m'. \quad (119)$$

There remains now the task of evaluating

$$\tau_{nm} = \int_0^\infty W_n(u) W_m(u) \frac{du}{u^3} - \int_0^{v_0} W_n(v) W_m(v) \frac{dv}{v^3}. \quad (120)$$

This expression involves integrals over the product $F_{nm}(u)$ of the functions $W_n(u)$ and $W_m(u)$

$$F_{nm}(u) \equiv W_n(u) W_m(u), \quad (121)$$

each of which satisfies a linear differential equation of the second order (72'). From the general theory of the products of solutions of linear differential equations of the second order⁹ it follows that

⁹ G. N. Watson, *Theory of Bessel Functions* (The Macmillan Company, New York, 1944), p. 146.

or

$$\frac{d}{du} \left\{ u^{\frac{1}{2}} \left(- \left[1 + \frac{1}{16u^2} + \frac{(\gamma_n + \gamma_m)}{2u} \right] W_n W_m + \left[\frac{1}{4u} + \gamma_n + \gamma_m \right] (\dot{W}_n W_m + W_n \dot{W}_m) - 2\gamma_n W_n \dot{W}_m - 2\gamma_m W_m \dot{W}_n - \dot{W}_n \dot{W}_m \right) \right\} = -[1 + (\gamma_n - \gamma_m)^2] W_n W_m / u^{\frac{1}{2}}. \quad (124)$$

Using (124) and the defining Eqs. (76), (77), and (83) we now deduce that

$$\int_0^\infty W_n(u) W_m(u) \frac{du}{u^{\frac{1}{2}}} = - \frac{[B_n B_m + 2(\gamma_n + \gamma_m) A_n A_m]}{4[1 + (\gamma_n - \gamma_m)^2]}, \quad (125)$$

$$\int_0^\infty W_n(v) W_m(v) \frac{dv}{v^{\frac{1}{2}}} = \frac{\{-[B_n B_m + 2(\gamma_n + \gamma_m) A_n A_m] + 4(v_0)^{\frac{1}{2}} \dot{W}_n(v_0) \dot{W}_m(v_0)\}}{4[1 + (\gamma_n - \gamma_m)^2]}, \quad (126)$$

$$A_n \equiv \frac{\pi^{\frac{1}{2}} 2^{1/4} e^{i\pi/8} C_n}{\Gamma(\frac{3}{4} - \kappa_n)}, \quad B_n \equiv - \frac{2(\pi)^{\frac{1}{2}} 2^{3/4} e^{i3\pi/8} C_n}{\Gamma(\frac{1}{4} - \kappa_n)}. \quad (127)$$

Adding (125) and (126) we finally get

$$\tau_{nm} = \int_0^\infty W_n(u) W_m(u) \frac{du}{u^{\frac{1}{2}}} - \int_0^{v_0} W_n(v) W_m(v) \frac{dv}{v^{\frac{1}{2}}} = - \frac{(v_0)^{\frac{1}{2}} \dot{W}_n(v_0) \dot{W}_m(v_0)}{[1 + (\gamma_n - \gamma_m)^2]}. \quad (128)$$

This completes the proof of Eq. (112), and with it of Eq. (86). Values of $d\gamma_m^0/dv_0$ are shown in Table IV.

5. ASYMPTOTIC SOLUTIONS FOR TRAPPED MODES

So far we have treated the asymptotic solutions for leaky modes and transitional modes, and have found that the formula derived for the characteristic values of transitional modes can be used with sufficient accuracy over a region extending into the leaky modes on one side and into the trapped modes on the other side. There remains now the treatment of completely trapped modes. The asymptotic theory of these modes has been developed by Furry⁴ along the lines of Gamow's treatment of the problem of radioactive decay in wave mechanics. Furry follows the technique of the W.K.B. method in which the height-gain functions are given different (asymptotic) representations in each of three regions of height, and the characteristic values are determined from Jeffreys'¹⁰ "connecting formulae." This method does not supply height-gain functions in two isolated regions centered around the two turning points. The principal aim of this section is to fill in these gaps by developing the

theory *ab initio* along Langer's method,⁵ in which the difficulty with the Stokes phenomena is completely obviated. The expression for the characteristic value we arrive at is, in the limiting case of completely trapped modes treated, identical with Furry's result. However, the method developed here can be applied also to cases of incomplete trapping when Furry's formula becomes inaccurate. The latter extension is reserved for another occasion.

Referring to Fig. 1, a trapped mode of order m is characterized by the condition that

$$F(h_0) = k \int_0^{h_0} [y(h) - y(h_0)]^{\frac{1}{2}} dh > v_m \simeq \pi(m - \frac{1}{2}). \quad (129)$$

The quantity in brackets in the differential equation

$$\frac{d^2 U_m(h)}{dh^2} + k^2 [\Lambda_m + y(h)] U_m(h) = 0 \quad (130)$$

vanishes at *two* turning points which, though complex, are extremely close to the real values

¹⁰ H. Jeffreys, Proc. Lond. Math. Soc. [2] 23, 428 (1923); Phil. Mag. 33, 451 (1943).

h_1 and h_2 shown in the figure. We shall use different representations of $U_m(h)$ in the regions I ($h < h_0$) and II ($h > h_0$). In region I let

$$u_1(h) = k \int_{h_1}^h [y(h) + \Lambda_m]^{1/4} dh \\ = e^{i\pi/2} k \int_{h_1}^h [-y(h) - \Lambda_m]^{1/4} dh, \quad 0 < h < h_1, \quad (131)$$

$$u_1(h) = e^{-i\pi/2} k \int_{h_1}^h [-y(h) - \Lambda_m]^{1/4} dh, \\ h_1 < h < h_0. \quad (132)$$

Here h_1 is strictly speaking a complex number which is the root of $[y(h_1) + \Lambda_m]$, but since the integrand vanishes in the neighborhood of h_1 , and $Im h_1$ is small, we may replace the lower limit by $Re h_1$; the same holds for h_2 below. In Langer's theory the leading term in the asymptotic expansion of (130) is according to Eqs. (12), (14), (16), and (20), given by

$$U_m^I(h) = \frac{A F^{(1)}(u_1)}{[y(h) + \Lambda_m]^{1/4}} + \frac{B F^{(2)}(u_1)}{[y(h) + \Lambda_m]^{1/4}}, \quad (133)$$

where

$$F^{(1)}(u_1) \equiv (u_1)^{1/2} II_{1/3}^{(1)}(u_1), \\ F^{(2)}(u_1) \equiv (u_1)^{1/2} II_{1/3}^{(2)}(u_1). \quad (134)$$

$$\frac{1}{U_m^I(h_0)} \frac{dU_m^I(h_0)}{dh} = \frac{1}{U_m^{II}(h_0)} \frac{dU_m^{II}(h_0)}{dh}, \quad (139)$$

or

$$\frac{II_{1/3}^{(2)}(u_0)}{II_{1/3}^{(1)}(u_0)} = \frac{[F^{(2)}(u_2) F^{(2)}(u_1) - F^{(2)}(u_2) F^{(2)}(u_1)]}{[F^{(2)}(u_2) F^{(1)}(u_1) - F^{(2)}(u_2) F^{(1)}(u_1)]}, \quad (140)$$

where now

$$u_1 = u_1(h_0) = e^{-i\pi/2} k \int_{h_1}^{h_2} [-y(h) - \Lambda_m]^{1/4} dh; \\ u_2 = u_2(h_0) = e^{-i\pi/2} k \int_{h_0}^{h_2} [-y(h) - \Lambda_m]^{1/4} dh. \quad (141)$$

Equation (140) serves to determine Λ_m , no matter what the magnitude of the arguments of

The constants A and B are subject to the boundary conditions $U_m(0) = 0$, from which it follows that

$$U_m^I(h) = \frac{C_m}{[y(h) + \Lambda_m]^{1/4}} [II_{1/3}^{(2)}(u_0) F^{(1)}(u_1) \\ - II_{1/3}^{(1)}(u_0) F^{(2)}(u_1)]; \quad u_0 \equiv u_1(0). \quad (135)$$

In region II let

$$u_2(h) = k \int_{h_2}^h [y(h) + \Lambda_m]^{1/4} dh, \quad h > h_2, \quad (136)$$

$$u_2(h) = e^{-i\pi/2} k \int_{h_2}^h [-y(h) - \Lambda_m]^{1/4} dh, \\ h_0 < h < h_2, \quad (137)$$

then a solution which represents an outgoing wave at great heights is

$$U_m^{II}(h) = \frac{E_m F^{(2)}(u_2)}{[y(h) + \Lambda_m]^{1/4}}. \quad (138)$$

The relation between E_m and C_m , and the value of Λ_m are determined from the required continuity of $U_m(h)$ and of $[dU_m(h)/dh]$ at $h = h_0$. The equation determining Λ_m is

the functions II and F may be, the only restriction being that the quantity A defined in (17) shall be small at both turning points so as to assure that the leading term in (20) is a sufficient approximation to $G(u)$. We shall however limit the discussion here to the case of highly trapped modes when the arguments of the functions in (140) are found to be large, thus allowing their representation by asymptotic expressions. In view of the phases of the u 's given in (141) the

appropriate asymptotic expressions are

$$F^{(2)}(u_1) \rightarrow \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp \left[i \left(-u_1 + \frac{5\pi}{12} \right) \right] P(u_1),$$

$$-2\pi < \arg u_1 < \pi, \quad (142)$$

$$\dot{F}^{(2)}(u_1) \rightarrow -i \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp \left[i \left(-u_1 + \frac{5\pi}{12} \right) \right] Q(u_1),$$

$$-2\pi < \arg u_1 < \pi, \quad (143)$$

$$F^{(1)}(u_1) \rightarrow \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp \left[i \left(u_1 - \frac{5\pi}{12} \right) \right] \bar{P}(u_1),$$

$$-\pi < \arg u_1 < 2\pi, \quad (144)$$

$$\dot{F}^{(1)}(u_1) \rightarrow i \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp \left[i \left(u_1 - \frac{5\pi}{12} \right) \right] \bar{Q}(u_1),$$

$$-\pi < \arg u_1 < 2\pi, \quad (145)$$

$$F^{(2)}(u_2) \rightarrow \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left[\exp \left[i \left(-u_2 + \frac{5\pi}{12} \right) \right] P(u_2) \right. \\ \left. + \frac{1}{2} \exp \left[i \left(u_2 - \frac{\pi}{12} \right) \right] \bar{P}(u_2) \right]^*,$$

$$-3\pi < \arg u_2 < \pi, \quad (146)$$

$$\dot{F}^{(2)}(u_2) \rightarrow -i \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left[\exp \left[i \left(-u_2 + \frac{5\pi}{12} \right) \right] Q(u_2) \right. \\ \left. - \frac{1}{2} \exp \left[i \left(u_2 - \frac{\pi}{12} \right) \right] \bar{Q}(u_2) \right],$$

$$-3\pi < \arg u_2 < \pi, \quad (147)$$

where

$$P(x) = 1 + \frac{5i}{72x} - \frac{385}{10368x^2} + \dots, \quad (148)$$

$$Q(x) = 1 + \frac{5i}{72x} + \frac{335}{10368x^2} + \dots$$

$$= P(x) + i\bar{P}(x), \quad (149)$$

and the bar denotes the complex conjugate

* The factor $\frac{1}{2}$ appearing in the brackets of (146) and (147) was introduced by Dr. W. F. Eberlein as leading to the correct Gamow formula. In textbooks two alternative formulae are given, with this factor having either a value of 1 or 0.

function. Using these expressions we find that

$$\frac{[F^{(2)}(u_2)\dot{F}^{(2)}(u_1) - \dot{F}^{(2)}(u_2)F^{(2)}(u_1)]}{[F^{(2)}(u_2)\dot{F}^{(1)}(u_1) - \dot{F}^{(2)}(u_2)F^{(1)}(u_1)]} \rightarrow \epsilon,$$

$$\epsilon = \frac{1}{2} \exp \left[i \left(-2u_1 + \frac{5\pi}{6} \right) \right] [P(u_1)Q(u_2) \\ - P(u_2)Q(u_1)] - \frac{1}{4} \exp \left[i \left(2u_2 - 2u_1 + \frac{\pi}{3} \right) \right] \\ \times [\bar{P}(u_2)Q(u_1) + \bar{Q}(u_2)P(u_1)]. \quad (150)$$

Here, the second term is of lower order of magnitude than the first, but it has been included because it alone contributes to the *imaginary part* of Λ_m , whereas the first term though larger, merely introduces a small change in the real part of Λ_m .

With regard to the left-hand member of Eq. (140) we set

$$u_0 = e^{i\tau} k \int_0^{h_1} [y(h) + \Lambda_m] dh = e^{i\tau} v_m + \Delta, \quad (151)$$

where v_m is a root of Eq. (22'). Assuming that Δ is small in the case of completely trapped modes we find from (140):

$$\frac{II_{1/3}^{(2)}(u_0)}{II_{1/3}^{(1)}(u_0)} \sim \frac{e^{i\tau/6} \Delta}{\pi v_m J_{1/3}^2(v_m)} = \epsilon, \quad (152)$$

$$\Delta = -\frac{1}{2} \pi v_m J_{1/3}^2(v_m) \exp(-i2u_1) \\ \times [P(u_1)Q(u_2) - P(u_2)Q(u_1)] \\ - \frac{1}{4} \pi v_m J_{1/3}^2(v_m) \exp \left[i \left(2u_2 - 2u_1 + \frac{\pi}{2} \right) \right] \\ \times [\bar{P}(u_2)Q(u_1) + \bar{Q}(u_2)P(u_1)]. \quad (153)$$

Since Λ_m is to a high approximation a purely real number, Eq. (141) shows that u_1 and u_2 are purely imaginary (negative and positive, respectively) numbers. From this fact it follows that the expressions in brackets in (153) are both real, the latter being approximately equal to 2. Hence

$$Im(\Delta) = -\frac{\pi}{2} v_m J_{1/3}^2(v_m) \exp[i(2u_2 - 2u_1)] \\ = -\frac{\pi}{2} v_m J_{1/3}^2(v_m) e^{-2\tau}, \quad (154)$$

$$\tau = k \int_{h_1}^{h_2} [-y(h) - \Lambda_m] dh. \quad (155)$$

In order to determine Λ_m we write

$$\Lambda_m = \Lambda_m^0 + i\delta \quad (156)$$

and obtain from (151)

$$k \int_0^{h_1} [y(h) + \Lambda_m]^{\frac{1}{2}} dh = k \int_0^{h_1} [y(h) + \Lambda_m^0]^{\frac{1}{2}} dh + \frac{1}{2} k \delta \theta i = v_m - \Delta, \quad (157)$$

$$k \int_0^{h_1} [y(h) + \Lambda_m^0]^{\frac{1}{2}} dh = v_m - \operatorname{Re}(\Delta), \quad (158)$$

$$\delta = -\frac{2\operatorname{Im}(\Delta)}{k\theta} = \frac{\pi v_m J_{1/2}(v_m) e^{-2r}}{k\theta} \sim \frac{e^{-2r}}{2k\theta};$$

$$\theta = \int_0^{h_1} \frac{dh}{[y(h) + \Lambda_m^0]^{\frac{1}{2}}}. \quad (159)$$

The last expression agrees with Furry's result obtained by the W.K.B. method.

Here we shall point out that although in the matter of the determination of characteristic values we have so far merely shown that our method leads to results agreeing with those obtained by the W.K.B. method for the limiting case of highly trapped modes, expression (140) can actually be used also to determine characteristic values for modes which are only moderately trapped by applying a technique developed by the writer for the bilinear model. These developments will, however, be reported in another communication.

The derivation of the height-gain functions as well as of the normalization factors can be carried out in the manner shown in previous sections. We shall therefore limit ourselves here to a statement of results for the case of *highly trapped modes*:

1. $0 \leq h \leq h_1$.

$$x = \int_h^{h_1} [y(h) + \Lambda_m^0]^{\frac{1}{2}} dh,$$

$$U_m(h) = \frac{x^{\frac{1}{2}} [J_{1/2}(x) + J_{-1/2}(x)]}{(u_m \theta)^{\frac{1}{2}} [y(h) + \Lambda_m^0]^{\frac{1}{4}}}. \quad (160)$$

2. $h_1 \leq h \leq h_0$.

$$y = k \int_{h_1}^h [-y(h) - \Lambda_m^0]^{\frac{1}{2}} dh, \quad (161)$$

$$U_m(h) = \frac{y^{\frac{1}{2}} [I_{-1/2}(y) - I_{1/2}(y)]}{[u_m \theta]^{\frac{1}{2}} [-y(h) - \Lambda_m^0]^{\frac{1}{4}}}.$$

3. $h_0 \leq h \leq h_2$.

$$y_0 = k \int_{h_1}^{h_0} [-y(h) - \Lambda_m^0]^{\frac{1}{2}} dh,$$

$$w_0 = k \int_{h_0}^{h_2} [-y(h) - \Lambda_m^0]^{\frac{1}{2}} dh,$$

$$w = k \int_h^{h_2} [-y(h) - \Lambda_m^0]^{\frac{1}{2}} dh, \quad (162)$$

$$U_m(h) = U_m(h_0) \left[\frac{-y(h_0) - \Lambda_m^0}{-y(h) - \Lambda_m^0} \right]^{\frac{1}{4}} \times \frac{w^{\frac{1}{2}} [I_{1/2}(w) + e^{-\pi/2} I_{-1/2}(w)]}{(w_0)^{\frac{1}{2}} [I_{1/2}(w_0) + e^{-\pi/2} I_{-1/2}(w_0)]}$$

4. $h \geq h_2$.

$$v = k \int_{h_1}^h [y(h) + \Lambda_m^0]^{\frac{1}{2}} dh,$$

$$U_m(h) = U_m(h_0) \left[\frac{-y(h_0) - \Lambda_m^0}{-y(h) - \Lambda_m^0} \right]^{\frac{1}{4}} \times \frac{v^{\frac{1}{2}} [-J_{1/2}(v) + e^{-\pi/2} J_{-1/2}(v)]}{w^{\frac{1}{2}} [I_{1/2}(w_0) + e^{-\pi/2} I_{-1/2}(w_0)]}. \quad (163)$$

Here, $u_m (\simeq 3/\pi)$ is defined in Eq. (34).

We note that Eq. (160) yields

$$\frac{dU_m(0)}{dh} = k \left(\frac{2}{\theta} \right)^{\frac{1}{2}} [y(0) + \Lambda_m^0]^{\frac{1}{4}}, \quad (164)$$

which agrees formally with Eq. (36) for the case of leaky modes.

Letters to the Editor

Concerning "Computer for Solving Linear Simultaneous Equations"

C. C. EAGLESFIELD

The Mullard Radio Valve Company, Ltd., Bournemouth, England

I HAVE read with great interest the paper by Berry, Wilcox, Rock, and Washburn¹ describing a computer for solving simultaneous equations.

I notice that the operation of this machine follows the classical method (named by Southwell the "relaxation method") of satisfying each equation in turn until a desired degree of approximation is reached.

I am wondering whether the authors considered a somewhat different method, hardly suited to arithmetical computation, but well suited to an electrical computer.

If the equations to be solved are

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \cdots + m_1 &= 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + m_2 &= 0, \\ \text{etc.,}\end{aligned}$$

and if for arbitrary values of the variables x_1, x_2 , etc.

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \cdots + m_1 &= d_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + m_2 &= d_2, \\ \text{etc.}\end{aligned}$$

Then of course the solution can be obtained by modifying x_1 , etc., so that $d_1 = d_2 = \cdots = 0$; the usual way of doing this is by adjusting x_1 until $d_1 = 0$, then adjusting x_2 until $d_2 = 0$, and so on.

The method I have in mind is to adjust x_1 until Σd^2 is reduced to a minimum, then adjust x_2 until Σd^2 is reduced to a (lower) minimum, and so on until a sufficient approximation is reached.

It is almost self-evident that this method is convergent, and very probable that it is more rapidly convergent than the line-by-line method. The latter is by no means always convergent, and in any case careful preparation of the relative size of the coefficients and the order of the equations is necessary. This preparation would not be necessary with the method I am suggesting.

In the machine described d_1 , etc., are obtained as a.c. voltages, so that it is a simple matter to get an indicating instrument responding to Σd^2 by applying each d to a separate square law rectifier and adding the rectified currents.

From the point of view of the practical operation of the computer, it is evident that other types of rectifier could conveniently be substituted for the square law rectifiers. The essential thing would seem to be that the indication should be independent of the signs of the d 's, and should be weighted in favor of the large d 's. Thus "average" rectifiers would seem to be quite suitable.

Another interesting possibility is to use "peak" rectifiers. If each d is applied to the anode of a separate diode and,

the cathodes being strapped, the circuit is completed by a single capacitance, then the capacitance will be charged to the peak value of the largest d . An indicator of this type would thus respond to the d which was numerically largest, instead of to a summation of the d 's.

The peak rectifier is probably the most convenient practically, followed by the average rectifier and the square law rectifier in that order.

While it is doubtful whether convergence could be formally proved for any but the square law rectifier, it seems very probable that both the average and the peak rectifier would give converging results.

The substitution of the over-all balance for the line-by-line balance modifies the design of the computer considerably, of course, but I do not think it brings in any great problems. It should give a computer which is easier and quicker to use, particularly with sets of equations which converge slowly, if at all, on the line-by-line computer.

July 24, 1946 (received July 29)

Thank you for allowing me to read the interesting letter from Mr. C. C. Eaglesfield.

Several months ago Dr. W. M. Bleakney of Lockheed Aircraft Corporation suggested use of the same mathematical method to me in a private communication. In fairness to Dr. Bleakney I should like to see a statement of this appear at the same time as the letter from Mr. Eaglesfield.

Clifford E. Berry
Consolidated Engineering Corporation
Pasadena, California

¹ J. App. Phys. 17, 262 (1946).

Preparing Pigments for the Electron Microscope

A. M. CRAVATH

Shell Development Company, Emeryville, California
September 16, 1946

ONE of the most satisfactory methods of preparing pigments for electron microscopy does not seem to have been reported. It is an adaptation of the well-known Green's method used in light microscopy. It produces better dispersions than the other methods tried by the author and is convenient and rapid. Its greatest advantage, however, in comparison with other methods in which the pigment is dispersed in the film forming solution is that it does not impair the definition. Presumably this is caused by the film being attached to the particle at a spot in the shadow instead of around the periphery as it is when cast on water.

The technique is not critical and is easily modified to suit the particular problem. One version is as follows:

Place 1 mm² of the pigment on a microscope slide. Add two drops of 4 percent parlodion in 70 percent amyl acetate plus 30 percent diisobutyl ketone. Rub smooth with a narrow spatula (made by filing and polishing $\frac{1}{8}$ " drill rod) which has a blade with slightly elliptical instead of rectangular cross section and therefore rubs over the sample instead of scraping it up. The slowness of evaporation of the diisobutyl ketone prolongs the period during which the mixture is highly viscous. Violent rubbing during this period is a very convenient way to concentrate a large amount of work on a small amount of material and thoroughly break up agglomerates if they are not too strongly cemented. If this viscous shearing work is too mild, a $\frac{1}{8}$ " glass rod instead of a spatula may be used. Rolling and rubbing the rod over the very small sample in the early fluid stage breaks up the hard lumps more completely and conveniently than use of a muller, for lumps are more often missed by the point contact of the muller than by the line contact of the rod. When the paste is nearly dry add four drops of butyl acetate and mix thoroughly until half of the solvent has evaporated. Touch the wet spatula or rod crosswise to a clean microscope slide 1 cm from the end. Place the dry part of the rod across the slide between the wet stripe and the end and with one quick light stroke smear the liquid along the slide. Agglomerates which may have reformed since the mixing are broken up by the violent shearing as the thin layer is spread out and they have no chance to reform because the film dries so rapidly. When the film is dry breathe on it and float it off on distilled water. The thickness will not be uniform. Place the specimen screens where the thickness appears to be correct.

Specimens in which the sample is not incorporated within the film may be prepared by a slight modification of the method. The sample is mixed in a volatile non-solvent and smeared over a slide previously coated with film by dipping in $\frac{1}{2}$ percent parlodion solution and drying. However, no advantages have been observed to justify this more difficult procedure.

Coordination Number in Solid Solutions— A Mathematical Paradox

DOUGLAS RENNIE HUDSON
University of Leeds, Leeds, England
July 24, 1946

DURING the last ten years the application of statistical methods to the theory of solubility has proceeded apace, and striking progress has been made, especially in Britain. To apply these theoretically interesting results usually requires a knowledge of the number of *A* molecules

surrounding a dissolved *B* molecule of different size or shape.

In investigating the number of metal atoms *X* surrounding another metal atom *Y* in solid solution, the simplest possible model is usually set up in which both

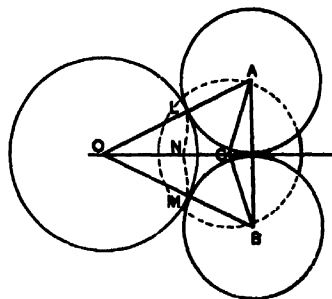


FIG. 1.

are assumed to be rigid spheres differing in diameter. In the course of calculation the following paradox has been discovered which may interest x-ray crystallographers and physicists generally.

If three spheres of radius *r* (atom *X*) with centers at *A*, *B*, *C*, are in contact with one another and with a large sphere of radius *R* (atom *Y*) with center at *O*, *ABC* is a plane equilateral triangle of side *2r*. Similarly the corresponding points of contact of the outer spheres with the kernel sphere, at *L*, *M*, *N*, form an equilateral spherical triangle on the surface of the latter. (Fig. 1.)

If now the radius of the larger sphere shrinks to *r*, with suitable accommodation of the smaller spheres, the lengths *OA*, *OB*, *OC*, will shrink to *2r* from *(R+r)*, giving four equal spheres in contact.

Then each side of the spherical triangle *LMN* will subtend an angle of 60° at the center *O*, or in mathematical notation each *side* of the triangle, *LM*, *MN*, or *NL*, is 60° or $\pi/3$. Then by the formula

$$\sin\left(\frac{\theta}{2}\right) = \frac{1}{2} \sec\left(\frac{\text{side } LM}{2}\right) = \frac{1.15470}{2} = .57735,$$

whence θ , the spherical angle, is found to be 70° 31.8', at *L*, *M*, and *N*. Since the spherical triangle *LMN* is equilateral, the spherical excess is $3 \times (70^\circ 31.8') - 180^\circ$, or 31° 35.4', or .55141 radians, and the area of the spherical triangle is .55141*r*². Now the total surface of the sphere is 4*πr*² and if this be divided by .55141*r*² the total number of such triangles which can be inscribed on the surface would appear to be (4*πr*²/.55141*r*²) = 22.8. But it is well known that in close packing, whether hexagonal or cubic, the number of spheres surrounding a given sphere is 12. Explanation of the paradox may entertain geometers and afford a half-hour's amusement to readers of mathematical bent.

Note on a Paper of H. Bateman

F. W. J. OLVER
Mathematics Division, National Physical Laboratory,
Teddington, England
September 30, 1946

IN a paper¹ entitled "Some Integral Equations of Potential Theory" the late Professor Bateman includes a table of $P_n(1-2e^{-t})$ to 15D for $n=1(1)10$, $t=1(1)20$. A number of errors were noticed in this table and subsequent

investigation showed the need for complete recomputation. This has been done and the new table, in which the values given are correct to within 0.52 units of the 15th decimal, is appended.

Attention is also drawn to the erroneous value given² for $P_4(\mu)$ where $\mu=1-2e^{-\pi}$; the last four decimals given should read ... 7662 instead of ... 5162.

¹ J. App. Phys. 17, 91-102 (1946).

² Reference 1, p. 99.

$$P_n(1-2e^{-t})$$

$n \backslash t$	1	2	3	4	5
1	+ 0.26424 11176 571.13	- 0.39526 49476 08978	- 0.35023 51643 18206	+ 0.13449 18185 76736	+ 0.34415 78146 74583
2	- 0.72932 94335 26775	- 0.27788 21339 12729	- 0.12412 25477 10654	- 0.38164 10835 56292	- 0.40197 48161 8827
3	- 0.90042 5832 64272	- 0.71615 1028 9890	- 0.4744 85488 0890	- 0.21049 90508 55291	- 0.03338 98178 26623
4	- 0.96336 8722 2252	- 0.89211 89424 35014	- 0.79015 33279 25199	- 0.66302 65464 98307	- 0.51760 76040 28999
5	- 0.98652 4190 01889	- 0.93984 47175 84022	- 0.82050 55158 57439	- 0.68928 43716 52798	- 0.57222 55850 05077
6	- 0.99504 24956 46667	- 0.98516 45222 14122	- 0.87023 59956 51009	- 0.75921 78000 33021	- 0.65281 72728 75236
7	- 0.99817 6250 68891	- 0.98453 56573 78988	- 0.89901 53471 08798	- 0.88183 70921 56184	- 0.75281 72728 75236
8	- 0.99932 90747 44195	- 0.98798 78594 43833	- 0.93597 78237 65885	- 0.92520 06735 53732	- 0.82655 97524 16840
9	- 0.99975 31803 91827	- 0.99925 96235 53558	- 0.98851 95392 12762	- 0.99153 31773 53226	- 0.98630 09051 20736
10	- 0.99990 32001 40475	- 0.99972 76127 88247	- 0.99945 52626 75897	- 0.99909 21868 95476	- 0.99863 83348 96966
11	- 0.99996 55965 98420	- 0.99989 91314 65238	- 0.99975 75978 73951	- 0.99956 59906 88745	- 0.99949 90075 52514
12	- 0.99998 77115 75293	- 0.99996 31148 52329	- 0.99992 52705 85236	- 0.99987 71131 50525	- 0.99981 54815 57023
13	- 0.99999 54793 41186	- 0.99998 61348 54253	- 0.99997 32627 87114	- 0.99995 47958 71677	- 0.99993 23911 80923
14	- 0.99999 83388 42562	- 0.99999 50108 31834	- 0.99999 02216 78114	- 0.99998 33624 67847	- 0.99997 50542 83623
15	- 0.99999 93911 95359	- 0.99999 81645 86658	- 0.99999 63291 74961	- 0.99999 38819 62012	- 0.99999 03239 50026
16	- 0.99999 97718 73258	- 0.99999 93247 80038	- 0.99999 86495 73251	- 0.99999 77492 97645	- 0.99999 68238 47018
17	- 0.99999 99172 01216	- 0.99999 97516 03717	- 0.99999 95032 05255	- 0.99999 91710 12610	- 0.99999 87580 19044
18	- 0.99999 99655 10914	- 0.99999 99086 20123	- 0.99999 98172 40250	- 0.99999 96954 00426	- 0.99999 95431 00356
19	- 0.99999 99837 94407	- 0.99999 99663 83222	- 0.99999 99327 66444	- 0.99999 98879 44074	- 0.99999 98319 16113
20	+ 0.99999 99958 77693	+ 0.99999 99876 33078	+ 0.99999 99752 66157	+ 0.99999 99587 76928	+ 0.99999 99381 65392

$n \backslash t$	6	7	8	9	10
1	+ 0.95464 00014 53797	- 0.26817 48073 26644	- 0.18068 47715 50211	+ 0.14819 38418 87370	+ 0.23701 82165 74609
2	- 0.21928 12946 65080	- 0.04754 01993 22603	- 0.25688 19941 08463	- 0.31162 84980 67465	- 0.20063 77689 23104
3	- 0.52378 90479 34200	- 0.36640 21347 36339	- 0.40985 37551 92309	- 0.37105 35279 20304	- 0.26611 15973 97119
4	- 0.56166 40008 03395	- 0.20339 42582 98305	- 0.50933 83202 15463	- 0.06810 25878 44205	- 0.20710 75153 98279
5	- 0.73556 57248 52066	- 0.65573 61016 67475	- 0.56931 56160 01227	+ 0.47800 57812 79961	- 0.83358 79750 75931
6	- 0.86244 13210 33978	- 0.86577 5712 05337	- 0.82913 12171 91742	- 0.78879 77202 45136	- 0.74507 07436 55250
7	- 0.93295 77118 18071	- 0.94856 90189 20862	- 0.93538 32422 93312	- 0.91956 30955 25272	- 0.90213 67456 08779
8	- 0.98482 31816 62874	- 0.98510 05568 44490	- 0.97598 81367 27241	- 0.97003 04866 29698	- 0.96343 20474 32366
9	- 0.99808 40684 77302	- 0.99745 91617 72468	- 0.99113 36665 21426	- 0.98982 32362 79290	- 0.98647 09001 53815
10	- 0.99928 8457 16543	- 0.99906 49156 19988	- 0.99573 38012 48168	- 0.99591 90856 76643	- 0.99501 32661 41742
11	- 0.99974 15589 33336	- 0.99965 59256 47297	- 0.99875 52622 70977	- 0.99849 73821 31571	- 0.99816 58427 25200
12	- 0.99996 50683 17055	- 0.99987 34244 15632	- 0.99954 32627 23045	- 0.99944 65024 60125	- 0.99925 13789 38820
13	- 0.99999 50760 84201	- 0.99995 34319 14456	- 0.99994 01368 04334	- 0.99992 51537 84321	- 0.99990 85338 04556
14	- 0.99999 71521 41841	- 0.99998 28635 40795	- 0.99997 12751 59630	- 0.99997 24899 76435	- 0.99996 63510 22665
15	- 0.99999 82735 27681	- 0.99999 36900 39790	- 0.99998 18974 83377	- 0.99998 88718 59320	- 0.99998 78211 68593
16	- 0.99999 82612 26878	- 0.99999 76816 36173	- 0.99998 70192 47002	- 0.99998 62240 59447	- 0.99998 54460 73600
17	- 0.99999 83603 40348	- 0.99999 91471 21310	- 0.99998 85034 41751	- 0.99998 83247 92187	- 0.99998 82479 02485
18	- 0.99999 97646 82563	- 0.99999 96862 43423	- 0.99999 95965 86666	- 0.99999 94557 48385	- 0.99999 93836 92145
19	- 0.99999 99134 31950	+ 0.99999 98845 75400	+ 0.99999 98515 86945	+ 0.99999 98144 96182	+ 0.99999 97732 73114

Here and There

New Appointments

J. H. Dillon, assistant director of the Firestone Research Laboratories, Akron, Ohio, and formerly chairman of the Division of High-Polymer Physics, American Physical Society, has been appointed director of research of the Textile Research Institute, with offices and laboratories at Princeton, New Jersey.

Leo Szilard has been appointed to the faculty of the University of Chicago, and will divide his time between the new Institute for Radiobiology and Biophysics and the division of social sciences.

Stanley S. Ballard joined the faculty of Tufts College, Medford, Massachusetts, this fall as professor of physics and chairman of the Department of Physics. His last previous academic association was at the University of Hawaii. During the war he served as a Commander in the United States Naval Reserve, and was a member of the technical staff, Joint Task Force One, during Operation Crossroads.

Professor G. H. Henderson of the Department of Physics, Dalhousie University, Halifax, Nova Scotia, has returned to his post after having been engaged for the past six years in research for the Royal Canadian Navy.

Varnakale L. Jones, formerly with the Stanolind Oil and Gas Company, has rejoined the engineering staff of The University of Tulsa as assistant professor of geophysics and head of the newly created Department of Geophysics.

Henry H. Young, Price D. Wickersham, Jr., William W. Niven, Jr., and George L. Tuer, Jr., have joined the technical staff of the Midwest Research Institute.

Albert M. Stone, formerly with the Radiation Laboratory, Massachusetts Institute of Technology, has received an appointment as a physicist with the Office of Naval Research. He will be stationed at its London office, serving in a liaison capacity between European and American physicists.

Lee A. DuBridge is the new president of California Institute of Technology, succeeding Dr. R. A. Millikan, who retired last spring. Dr. DuBridge was formerly with the Department of Physics at the University of Rochester and later at the NDRC Radiation Laboratory, Cambridge, Massachusetts.

William A. Lewis, Jr., formerly research professor of electrical engineering at Illinois Institute of Technology, became dean of the Institute's graduate school on September 1.

Honors and Awards

Raymond D. Mindlin, associate professor of civil engineering at Columbia University, was recently awarded the Medal for Merit for his wartime work in the development of the radio proximity fuse.

In recognition of war research on fire-control instruments for better marksmanship in shooting down enemy aircraft, 27-year-old Wayne G. Norton of the Eastman Kodak Company received the Adolph Lomb medal from the Optical Society of America on October 4. The award is made annually "to a person under 30 years of age who shall have made a noteworthy contribution to optics."

Eugene C. Crittenden, associate director of the National Bureau of Standards, has been elected to membership in the International Committee on Weights and Measures. Commented Dr. E. U. Condon, director of the Bureau: "Dr. Crittenden's election is not only a distinct personal honor, but is recognition of the leading role played by the United States through the Bureau of Standards in the field of measurements." Dr. Crittenden is an authority on the measurement of light.

New Division at Brown University

Science reports that a new graduate division of applied mathematics has been organized at Brown University as an interdepartmental project of the Departments of Mathematics and Physics and the Division of Engineering, to meet the needs of schools and research laboratories for engineers, physicists, and mathematicians whose training extends beyond the accepted boundaries of their respective fields. The following professors have been appointed to the new division: William Prager (chairman), applied mechanics; John Henry Marchant (director of research), engineering; Maurice Anthony Biot, applied physical sciences; George Frances Carrier, engineering; James Arthur Krumhansl, physics; Chia Chiao Lin, applied mathematics; and Rohn Truell, physics.

Students are admitted who have excellent undergraduate records in two of the following subjects: engineering, physics, and mathematics. Training in research is stressed for the advanced students. Fellowship funds have been provided by the Rockefeller Foundation.

Radio Physics Department at Tufts College

On September 1 a Department of Radio Physics was established at Tufts College, with Professor Jamison R. Harrison appointed as head. Work in the department will comprise the field of radio and the allied fields of acoustics and electromagnetics. While the new department will be administered independently of the department of physics, the two departments will operate on a close cooperative basis. A broad foundation in physics will be required of majors in radio physics, and the urgent need on the part of majors in physics today for basic knowledge of electronics will be met by courses in radio physics.

The following personnel are associated with Professor Harrison in the new department while serving also in the

department of physics: Charles R. Mingins, associate professor of physics and radio communication; Carl A. Stevens, assistant professor of physics and radio communication; Domina Eberle Spencer, assistant professor of physics.

Information Clearing House

A library research service of special interest to industrial and scientific organizations not having ready access to the wealth of scientific literature in New York City is now offered by the Technical Library Consultants, 347 Madison Avenue, New York 17. The service is specifically designed to meet the literature research needs of executive, sales, and research departments.

Humble Oil Lectures

Establishment of an annual series of scientific lectures to be given at its Baytown Refinery and to be known as "Humble Lectures in Science" has been announced by Humble Oil and Refining Company. The series is designed especially to keep Humble technical personnel abreast of the latest developments in science and engineering. Outstanding scientists will be invited each year to present topics in their fields of special study and research.

Pittsburgh Physical Society Officers

Following are the 1946-47 officers of the Physical Society of Pittsburgh: president, Dr. I. Estermann, Carnegie Institute of Technology; vice president, Dr. Frank Morgan, Gulf Research and Development Corporation; secretary-treasurer, Dr. E. Stickley, Pittsburgh Plate Glass Company.

Atomic Research Center on Long Island

Nine leading educational institutions of the East have cooperated to form Associated Universities, Inc., which will operate as contractor with the government a large atomic research center at Camp Upton, Long Island. The camp has been transferred from surplus to the Manhattan Engineer District. Director of scientific research for Associated Universities will be Dr. Philip M. Morse, professor of physics at Massachusetts Institute of Technology and wartime consultant on scientific projects to the Navy Department.

The primary purpose of the project will be one of pure research, looking toward the peacetime application of atomic energy in industrial and medical fields and toward the training of more experts in the field of atomic research. Design of machines, which will cost approximately \$5,000,000, will start immediately. Research, according to Dr. Morse, will be under way by early spring.

New Curriculum at Brooklyn Polytechnic

A new curriculum leading to the degree of Master of Science in Applied Mechanics has been announced by the Polytechnic Institute of Brooklyn, "to make available to graduate engineers a broad knowledge of the principles of mechanics which became so all-important in the new concepts of design developed during the war." These

studies are incorporated in the new Department of Aeronautical Engineering and Applied Mechanics headed by Professor R. P. Harrington.

Universal Oil Products Company Expansion

The physics division of Universal Oil Products Company has recently moved into a new home, a functional building designed largely by the scientists who are to work in it. The building is completely air conditioned, and all the equipment is mounted on steel supports resting on spread concrete footings and insulated from the floor by cork gaskets. The laboratory is located at Riverside, Illinois. Dr. G. M. Webb is director of the physics division.

Kimberly-Clark Fellowships

The Kimberly-Clark Corporation, Neenah, Wisconsin, has announced a program of fellowships, with grants offered to the University of Wisconsin, the University of Minnesota, and the Massachusetts Institute of Technology. The fellowships are in the fields of physics, physical chemistry, and chemical engineering.

Louisville Physics Society

The Louisville Physics Society ended its first year with a dinner meeting on June 6. The new officers for the coming year were installed: Dr. Peter L. Vissat, president, and Dr. H. T. Smyth, secretary-treasurer.

Pittsburgh Spectroscopists

The Spectroscopy Society of Pittsburgh was organized in August. It will cooperate in the sponsoring of an Applied Spectroscopy Conference to be held at the University of Pittsburgh November 15, 1946. For information about the Society address Dr. Mary E. Warga, University of Pittsburgh, Pittsburgh 13, Pennsylvania.

Westinghouse Program for Fundamental Research

Westinghouse Research Laboratories have announced a broad program of research into nuclear physics and the problems associated with the conversion of atomic energy into a useful source of power for the peacetime world. The program will be in direct charge of Dr. W. E. Shoupp, Manager of the Electronics Department. First step will be modernization of the atom-smasher at the East Pittsburgh Laboratory.

Radioisotopes from the Manhattan Project

A detailed announcement on the availability and procurement of pile-produced radioisotopes from the Manhattan Project appeared in the June 14 issue of *Science*. Tables are included giving pertinent data on the characteristics and the quantities which may be made available of around 100 isotopes and isotopic mixtures. For practical reasons isotopes with a half-life less than 12 hours are not considered for distribution. Most of the isotopes are produced by fission or (n, γ) processes. Only four isotopes are produced by the (n, p) process with sufficient yield for distribution. Other processes are either not sufficiently productive or do not occur.

The article emphasized that (1) present piles were not designed for tracer and therapeutic isotope production, (2) waste plutonium process solutions are not a feasible source for separated fission isotopes, (3) routine production methods and facilities are not yet developed for most isotopes, (4) isotopes which can now be made available are only experimental lots resulting from research and development proceedings, (5) technical problems involved in the irradiation and processing of essential materials has been and will continue to be responsible for the delay in making certain isotopes available by routine production.

Allocation and distribution will be effected on the basis of the general policies, as well as on recommendations regarding specific applications, made by well qualified advisory groups nominated for Manhattan District appointment by the National Academy of Sciences. Charges will be made for materials and services on the basis of "out-of-pocket" operational expenses to the government necessitated by the non-project production and service program. Costs for construction or rental of major plant facilities and the expenses of research and development on isotope production will be assumed by the Project.

All correspondence concerning radioisotope procurement should be addressed to the Isotopes Branch, Research Division, Manhattan District, P.O. Box E, Oak Ridge, Tennessee. Reference to the original article for pertinent details is recommended before instituting inquiries or requests.

Bulletin of Mathematical Biophysics

Following is the table of contents for the December 1946 issue of *The Bulletin of Mathematical Biophysics*:

- The Half-Life of a Drug in Relation to its Therapeutic Index—H. D. LANDAHL
Contributions to the Mathematical Theory of the Stability of Metabolizing Systems with Variable Surface Tension—ALFONSO SHIMMFL
Asynchrony of Labirinthine Receptors as a Physical Factor in Motion Sickness—MANUEL F. MORALFS
A Mathematical Description of Metabolizing Systems: I—HERMAN BRANSON

The University of Chicago Press, Chicago, Illinois.
Volume 8, Number 4.

Placement Register

In response to many requests from physicists and employers of physicists, the American Institute of Physics will operate another one of its Placement Registers in connection with the joint meeting of the American Physical Society and the American Association of Physics Teachers on January 30 and 31, and February 1, 1947.

The Register affords an effective supplement to the continuing Placement Service carried on by the Institute. The continuing service provides an opportunity for those desiring new employment to place their records on file at the Institute office. Here they may be consulted by employers or sent out to employers for consideration.

The Register itself provides an on-the-spot mechanism for bringing applicants and employers' representatives together. One or more rooms will be reserved for the Regis-

ter and interviews will be scheduled as requested. Because employers will have opportunity to consult the records of applicants before the interview period begins, pre-registration on the part of the applicants is essential for the maximum effectiveness of the Register.

Registration forms and further information may be obtained from the Institute, 57 East 55 Street, New York 22, New York.

Addendum

There should have been a footnote to the title of E. P. Wigner's paper on "Theoretical Physics in the Metallurgical Laboratory of Chicago," November, 1946, *Journal of Applied Physics*, which should read as follows:

This document is based on work performed under Contract No. W-7401-eng-37 for the Manhattan Project of the Metallurgical Laboratory at the University of Chicago.

New Books

Piezoelectricity

By WALTER G. CADY. Pp. 806+xxviii, Figs. 166, 6×9 in. International Series in Pure and Applied Physics. McGraw-Hill Book Company, New York, 1946. Price \$9.00.

The purpose and scope of this book is compactly stated in the sub-title: An Introduction to the Theory and Applications of Electromechanical Phenomena in Crystals. Because of the increasing importance of this subject and many applications, Professor Cady's book is a needed addition to the literature in this field, much of which has been heretofore scattered in scientific and technical journals in several languages.

A thorough study of piezoelectricity, such as this, must necessarily draw on much of the material of crystal physics. This is because piezoelectricity is closely related to the elastic and dielectric properties of crystals and less closely to pyroelectric and to piezo- and electro-optical effects. The Weiss theory of ferromagnetism which finds a close analogy in the theory of the Seignette-electric crystals is also discussed. There is an extensive treatment of resonator theory which makes full use of the electro-mechanical analogy and explains how the equivalent electrical constants of vibrating crystals may be computed over any range of frequencies. Details of electrical circuits, however, are generally omitted since they may be found in other books. In like manner, while there is a good chapter on practical applications, only the general principles are discussed, the details being left to an extensive list of classified references. In view of the wide use of oblique crystal cuts, the great number of specialized

formulae for calculating crystal constants for rotated axes will be found very useful.

Since the book is intended for research workers as well as for students of physics and for radio amateurs, parts of it will be found difficult or incomprehensible to many in the latter group. In order to fully understand all of this book, familiarity with theoretical physics and circuit theory is essential, though many parts may be read profitably with less background.

A commendable feature is the list of symbols and abbreviations giving the number of the section in which each quantity is first introduced. The reviewer, however, found it annoyingly inconvenient to look back for equations referred to only by number, there being an extremely large number of such cross references. It would appear that page references along with the numbers would have made this task much easier. There is an extensive bibliography at the end of the book. The tables of numerical constants, which are indexed for ready reference, will be found very useful to workers in this field. Some of the data included have not been published elsewhere, being taken from investigations carried out at Wesleyan University by students of Professor Cady and Van Dyke. Gaps in existing data are pointed out, and suggestions for further research to settle controversial topics may be found in many places.

Besides the few misprints, there is not much fault that one may find with this book. However, in defining a Seignette-electric crystal it is asserted that the possession of a Curie point on one side of which the dielectric properties show non-linearity and hysteresis is the basic criterion. Shortly thereafter one finds the statement that perhaps some day a crystal will be discovered which is stable only in the Seignette-electric range so that no Curie point is observable. Surely then, the existence of a Curie point cannot be regarded as basic a criterion as the non-linearity and hysteresis themselves.

Furthermore, it seems contradictory to attribute practically all hysteresis effects in Rochelle salt to the clamped dielectric properties (p. 479), suggesting that a free crystal would not show such effects, and then to state that the constraints imposed by cement; electrodes, clamping, and water-proofing greatly reduce the effects of temperature and hysteresis (p. 670). It is well known that the anomalies of Rochelle salt were discovered in crystals which were not clamped except for the constraint imposed by the electrodes and that in many cases this was quite negligible.

The use of literary quotations at the head of each chapter add interest to the book, giving it a certain cultural seasoning. Many of the quotations aptly express some central idea in the chapter, but some of them are so irrelevant that they would have been better omitted.

The wide variety of subjects treated testifies to Professor Cady's broad, scholarly interests, and thorough knowledge of his specialty. This book is a distinguished addition to the growing number of important volumes in the International Series in Pure and Applied Physics.

JOSEPH VALASEK
University of Minnesota

New Booklets

Mathematical Tables Project of the National Bureau of Standards, 150 Nassau Street, New York, New York, has published three short notes bearing on tables now in progress, as follows: Mathieu functions, exponential integrals for complex arguments, and functions related to certain confluent hypergeometric functions.

The Ohio Valley Spectrographic Society, Engineers Club Building, Dayton, Ohio, is distributing *Collected Abstracts, Published during 1945 on Spectrochemical Analysis*, compiled by Edwin S. Hodge. 64 pages, well indexed. \$1 postpaid for cash orders, \$1.25 when billing is required. Discount of 15 percent on lots of 10 or more copies.

North American Philips Company, Inc., 100 East 42 Street, New York 17, New York, has issued a new 8-page booklet (R1041) titled *The Geiger-Mueller X-Ray Spectrometer*. It explains the history and construction details covering the instrument, along with circuits and data pertinent to correct usage. Free.

General Electric Company, Schenectady, New York, has assembled in booklet form six addresses on *Applications of Atomic Power* given by staff members of the General Electric Research Laboratory on the Science Forum Program, Station WGY, Schenectady, during the fall of 1945. 26 pages.

Ampro Corporation, 2835 North Western Avenue, Chicago, Illinois, has issued a new price sheet, showing the general trend in prices of its 16-mm motion picture equipment and accessories to be slightly upwards. 4 pages.

Reiner Electronics Company, Inc., 152 West 25 Street, New York 1, New York, has published a new 4-page bulletin (105) providing a complete description of the new laboratory instruments now being produced by the company. The design of these volt-ammeters makes available for the first time a method of manual insertion of accurately calibrated shunts and multipliers with values always visible in the unit. This self-instructing feature teaches the student as he works. The principles of operation are thoroughly described and all applications and specifications are given in detail.

Interchemical Review for Spring 1946, published by the Research Laboratories of Interchemical Corporation, 432 West 45 Street, New York 19, New York, features the following articles: "Synthetic Amino Acids" and "A Brief Review of Atomic Physics." 28 pages.

Carson Micrometer Corporation, 28 Edison Place, Newark 2, New Jersey, has announced production of four redesigned models of its electronic micrometers for precise thickness measurement. They are described in a 4-page leaflet. Address Department 72 of the company.

Author Index to Volume 17

References with (L) are to Letters to the Editor.

- Adams, H. E. and P. O. Powers.** Thermal polymerization of drying oils—325
- Akeley, E. S.** Study of a certain type of resonant cavity and its application to a charged particle accelerator—1056
- Albright, John G.** Book review—537
- van Amerongen, G. J.** Permeability of different rubbers to gases and its relation to diffusivity and solubility—972
- Andrews, R. D., A. V. Tobolsky, and E. E. Hanson.** Theory of permanent set at elevated temperatures in natural and synthetic rubber vulcanizates—352
- Anthony, Robert L.** (see Guth, Eugene) 347
- Arnold, Orlan M.** (see Rathbun, Mary E.) 759
- Baker, R. F.** (see Hillier, James) —12
- Baker, T. C. and F. W. Preston.** Effect of water on the strength of glass 179
— Fatigue of glass under static loads —170
— Wide range static strength testing apparatus for glass rods—162
- Bargmann, V. and H. F. Ludloff.** Elastic limit for dynamic loading—63 (L)
- Barnes, R. Bowling, Robert S. McDonald, and Van Zandt Williams.** Erratum: Small prism infra-red spectroscopy—532
- Barry, Arthur J.** Viscometric investigation of dimethylsiloxane polymers—1020
- Bateman, H.** Some integral equations of potential theory—91
- Baum, Richard F.** Design of broad band I. F. amplifiers—519
— Part II—721
- Beams, J. W., J. L. Young, III, and J. W. Moore.** Production of high centrifugal fields—886
- Beck, Hugo M.** (see Faust, William R.) —749
- Bekkedahl, Norman** (see Wood, Lawrence A.) —362
- Berry, Clifford E.** (see Eaglesfield, C. C.) —1125 (L)
—, **Doyle E. Wilcox, Sibyl M. Rock, and H. W. Washburn.** Computer for solving linear simultaneous equations—262
- Birks, L. S. and H. Friedman.** Particle size determination from x-ray line broadening—687
- Blewett, John P.** Oxide coated cathode literature, 1940–1945—643
- Bloch, F., K. M. Hamermesh, and M. Phillips.** Radar reflections from long conductors—1015
- Boyer, R. F.** (see Spencer, R. S.)—398
- Brentano, J. C. M.** Para-focusing properties of microcrystalline powder layers in x-ray diffraction applied to the design of x-ray goniometers—420
- Bridgman, P. W.** Effect of hydrostatic pressure on plastic flow under shearing stress—692
— Studies of plastic flow of steel, especially in two-dimensional compression—225
— Tensile properties of several special steels and certain other materials under pressure—201
- Brittin, Wesley E.** Liquid rise in a capillary tube—37
- Brown, Gordon P., Albert DiNardo, George K. Cheng, and Thomas K. Sherwood.** Flow of gases in pipes at low pressures—802
- Bruck, George G.** Letter to the editor—213
- Bryant, J. M. and M. Newman.** Lightning recorded by aeroplane propeller—555
- Buerger, M. J.** Interpretation of Harker syntheses—579
— and **Gilbert E. Klein.** Correction of diffraction amplitudes for Lorentz and polarization factors—285
Erratum—743
- Burlington, Richard S.** Book review—409
- Busse, W. F., J. M. Lambert, and R. B. Verdery.** Tackiness of GR-S and other elastomers—376
- Carter, J. L.** (see Osborne, M. F. M.)—871
- Champayne, E. F.** (see McLachlan, Dan, Jr.) 1006
- Cheng, George K.** (see Brown, Gordon P.)—802
- Chilton, Ernest G.** Dynamic characteristics of rubber supports from vibration table data 492
- Coggeshall, Norman D. and Eleanor L. Saier.** Analyses of mixtures of light gases by infra-red absorption 450
- Coomes, Edward A.** Pulsed properties of oxide cathodes—647
- Cravath, A. M., A. E. Smith, J. R. Vinograd, and J. N. Wilson.** Preparation of electron microscope specimens for determination of particle size distribution in aqueous suspensions 309
- Cravath, Austin M.** Preparing pigments for the electron microscope—1125 (L)
- Crittenden, Eugene C., Jr.** (see Parkins, William E.)—447
— and **William E. Parkins.** Methods for betatron or synchrotron beam removal —444
- Dart, S. L. and Eugene Guth.** Elastic properties of cork. I. Stress relaxation of compressed cork—314
- Debye, P. P.** Photoelectric instrument for light scattering measurements and a differential refractometer—392
- Dienes, G. J. and H. F. Klemm.** Theory and application of the parallel plate plastometer—458
- DiNardo, Albert.** (see Brown, Gordon P.) —802
- Eaglesfield, C. C.; Clifford E. Berry.** Concerning "computer for solving linear simultaneous equations"—1125 (L)
- Eastwood, Mary J.** (see Rathbun, Mary E.)—759
- Eisenstein, A.** Study of oxide cathodes by x-ray diffraction methods. Part I. Methods, conversion studies, and thermal expansion coefficients—434
— Part II. Oxide coating composition—654
— X-ray method for measuring the thickness of thin crystalline films—874
— (see Fineman, A.)—663
- Ewell, Arthur W.** Recent ozone investigations—908

- Faust, William R. and Hugo M. Beck.** Oscillation conditions in single tuned amplifiers—749
- Feenberg, Eugene.** Elementary treatment of longitudinal debunching in a velocity modulation system—852
— Relation between nodal positions and standing wave ratio in a composite transmission system—530
— and **David Feldman.** Theory of small signal bunching in a parallel electron beam of rectangular cross section—1025
- Feldman, David.** (see Feenberg, Eugene)—1025
- Field, J. E., D. E. Woodford, and S. D. Gehman.** Applications of infra-red methods in the structural examination of synthetic rubber—386
- Field, Robert F.** Formation of ionized water films on dielectrics under conditions of high humidity—318
- Fineman, A. and A. Eisenstein.** Studies of the interface of oxide coated cathodes—663
- Fisher, J. R.** (see Teal, G. K.) 879
- Forsythe, George E.** Book review—535
- Friedman, H.** (see Birks, L. S.) 687
- Frocht, Max M.** Numerical solution of Laplace's equation in composite rectangular areas—730
- Froelich, H. C.** Depreciation of fluorescent lamps without passage of current—573
- Fry, D. L., R. E. Nusbaum, and H. M. Randall.** Analysis of multicomponent mixtures of hydrocarbons in the liquid phase by means of infra-red absorption spectroscopy—150
- Fuertes, Fidel Alsina.** Letter to the editor on the "power function"—712 (L)
- Fulcher, Gordon S.** Interpolation with the aid of a plot of first differences 617
- Gehman, S. D.** (see Field, J. E.)—386
- Gemant, Andrew.** Induction heating of hollow metallic cylinders—195
— Transient temperatures around heating pipes maintained at constant temperature—1076
- Glahtart, J. L.** (see Vonnegut, B.)—1082
— and **F. W. Preston.** Fatigue modulus of glass—189
- Gregory, J. H.** (see Smith, Alva W.)—33
- Guth, Eugene** (see Dart, S. L.)—314
—, **Paul E. Wack, and Robert L. Anthony.** Significance of the equation of state for rubber—347
- Hall, C. E.** Book review—538
- Hammermesh, K. M.** (see Bloch, F.)—1015
- Hansen, W. W.** Erratum: On maximum gain-band width product in amplifiers—109
- Hanson, E. E.** (see Andrews, R. D.)—352
— (see Kelsey, R. H.)—675
- Harris, Louis.** Evaporation of antimony—757 (L)
- Harvey, E. Newton** (see McMillen, J. Howard)—541
- Havenhill, R. S.** Book review—539
—, **H. C. O'Brien, and J. J. Rankin.** Electrostatic and tensile properties of rubber and GR-S at elevated temperatures—338
- Heidenreich, R. D., L. Sturkey, and H. L. Woods.** Investigation of secondary phases in alloys by electron diffraction and the electron microscope—127
- Heisler, Michael P.** (see Paschkis, Victor)—246
- Herman, Robert C. and Charles F. Meyer.** Thermoluminescence and conductivity of phosphors—743
- Hershbarger, W. D.** Absorption of microwaves by gases—495
— (see Walter, John E.)—814
- Hillier, James.** Further improvement in the resolving power of the electron microscope—307
— Study of distortion in electron microscope projection lenses—411
— and **R. F. Baker.** On the improvement of resolution in electron diffraction cameras—12
- von Hippel, A., J. H. Schulman, and E. S. Rittner.** New electrolytic selenium photo-cell—315
— (see Roberts, S.)—610
- Hollomon, J. H.** (see Zener, Clarence) -69
— and **Clarence Zener.** Problems in fracture of metals—82
- Honold, Edith** (see Wakeham, Helmut)—698
- Horvath, William J.** (see Iskenderian, Haig P.)—255
- Howlett, J.** Film-lubrication between spherical surfaces: with an application to the theory of the four-ball lubricant testing instrument—137
- Hudson, Douglas Rennie.** Coordination number in solid solutions—a mathematical paradox—1126 (L)
- Hull, Albert W.** Stresses in cylindrical glass-metal seals with glass inside—685
- Hull, Gordon Ferrie, Jr.** Resonant circuit modulator for broad band acoustic measurements—1066
- Hunter, Lloyd P.** Energy build-up in magnetrons—833
- Hurd, J. M.** (see Smoluchowski, R.)—864
- Ingersoll, H. G.** Fine structure of viscose rayon—924
- Iskenderian, Haig P. and William J. Horvath.** Determination of the temperature rise and the maximum safe current through multiconductor electric cables—255
- Jacobs, Harold.** Dissociation energies of surface films of various oxides as determined by emission measurements of oxide coated cathodes—596
- Johnson, J. B.** Cathode-ray tube for viewing continuous patterns—891
- Kaufmann, K.** (see Watson, John H. L.)—996
- Kelsey, R. H. and E. E. Hanson.** Method for preparing rubber latex specimens for the electron microscope—675
- King, Allen L.** Pressure-volume relation for cylindrical tubes with elastomeric walls: the human aorta—501
- King, D. D.** Measured impedance of cylindrical dipoles—844
- King, Ronald** (see Middleton, David)—273
- Kingdon, K. H.** (see Lafferty, J. M.)—894
- Klein, Gilbert E.** (see Buerger, M. J.)—285, 743
- Klemm, H. F.** (see Dienes, G. J.)—458
- Korff, Serge A.** Book review—408
- Krzywoblocki, M. Zbigniew.** Approximate calculation of the range of a winged jet projectile—515
- Lafferty, J. M.** Millimeter-wave reflex oscillator—1061
— and **K. H. Kingdon.** Improvements in the stability of the FP-54 electrometer tube—894

- Lambert, J. M.** (see Busse, W. F.)—376
- Landsberg, Peter T.** On plane rotations in n dimensions—60 (L)
- Langmuir, D. B.** (see Malter, L.)—456
- Liebhaufsky, Herman A.** Picturesque oxidation of molybdenum vacuum-tube leads—901
- Lucht, C. M.** (see Smoluchowski, R.) 864
- Ludeke, Carl A.** Experimental investigation of forced vibrations in a mechanical system not having a non-linear restoring force—603
- Ludloff, H. F.** (see Bargmann, V.)—63 (L)
- Lynn, J. T.** (see Smith, Alva W.)—33
- Lyons, W. James.** General relations of flow in solids and their application to the plastic behavior of tire cords—472
- Mack, Charles.** Plastic flow, creep, and stress relaxation. Part I—Plastic flow—1086
— Plastic flow, creep, and stress relaxation. Part II—Creep—1093
— Plastic flow, creep, and stress relaxation. Part III—Creep and elastic after-effect—1101
- MacLean, William R.** Microwave dielectric loss measuring technique—558
- Makinson, K. Rachel** (see Makinson, R. E. B.) -567
- Makinson, R. E. B., J. M. Somerville, K. Rachel Makinson, and P. Thonemann.** Magnetically-controlled gas discharge tubes -567
- Malter, L. and D. B. Langmuir.** Methods of minimizing lead loss in emissivity and resistivity determinations -456
- Marcum, Jess.** Interchange of energy between an electron beam and an oscillating electric field—4
- Margenau, Henry.** Book review—641
- Mayer, Ernest.** Resonant frequencies of the nose-in cavity—1046
- McCuskey, S. W.** Book review—536
- McDonald, Robert S.** (see Barnes, R. Bowling)—532
- McLachlan, Dan, Jr. and E. F. Champayne.** Machine for the application of sand in making Fourier projections of crystal structures—1006
- McMillen, J. Howard and E. Newton Harvey.** Spark shadowgraphic study of body waves in water—541
- McReynolds, Andrew W.** Electrical observations of the Austenite-Martensite transformation in steel—823
- Meyer, Charles F.** (see Herman, Robert C.)—743
- Middleton, David.** Response of biased, saturated linear, and quadratic rectifiers to random noise—778
— (see Van Vleck, J. H.)—940
— and **Ronald King.** Thin cylindrical antenna: a comparison of theories—273
- Moop, Parry and Domina Eberle Spencer.** Approximations to Planckian distributions—506
- Moore, J. W.** (see Beams, J. W.)—886
- Morris, M.** Book review—312
- Mueller, C. W.** Erratum: Secondary electron emission of Pyrex glass—62
- Newman, M.** (see Bryant, J. M.)—555
- North, H. Q.** Properties of welded contact germanium rectifiers—912
- Nusbaum, R. E.** (see Fry, D. L.)—150
- O'Brien, H. C.** (see Havenhill, R. S.)—338
- Olver, F. W. J.** Note on a paper of H. Bateman—1127 (L)
- Osborne, M. F. M. and J. L. Carter.** Transient analysis of linear systems, using underwater explosion waves—871
- Osgood, Thomas H.** Book review—540
- Parkins, William E.** (see Crittenden, Eugene C., Jr.)—444
— and **Eugene C. Crittenden, Jr.** Graphical method for determining particle trajectories—447
- Paschke, Victor and Michael P. Heisler.** Accuracy of lumping in an electric circuit representing heat flow in cylindrical and spherical bodies—246
- Pekeris, C. L.** Asymptotic solutions for the normal modes in the theory of microwave propagation—1108
— Perturbation theory of the normal modes for an exponential M-curve in non-standard propagation of microwaves—678
- Phillips, M.** (see Bloch, F.)—1015
- Pickett, Gerald.** Letter to the editor—61
- Powers, P. O.** (see Adams, H. E.)—325
- Preston, F. W.** (see Baker, T. C.)—162, 170, 179
— (see Glathart, J. L.)—189
- Randall, H. M.** (see Fry, D. L.)—150
- Rankin, J. J.** (see Havenhill, R. S.)—338
- Rathbun, Mary E., Mary J. Eastwood, and Orlan M. Arnold.** Supplementary bibliography of electron microscopy—759
- Read, Thomas A. and E. P. T. Tyndall.** Internal friction and plastic extension of zinc single crystals—713
- Reissner, Eric.** On vibrations of shallow spherical shells—1038
- Rittner, E. S.** (see Von Hippel, A.)—215
- Roberts, S. and A. Von Hippel.** New method for measuring dielectric constant and loss in the range of centimeter waves—610
- Robinson, A. L.** Book review—537
- Rock, Sibyl M.** (see Berry, Clifford E.)—262
- Rosenthal, Jenny E.** On a novel form of refrigerator—62 (L)
- Roseveare, W. E.** (see Waller, R. C.)—482
- Saier, Eleanor L.** (see Coggeshall, Norman D.)—450
- Sard, R. D.** Calculated frequency spectrum of the shot noise from a photo-multiplier tube—768
- Schroeder, Alfred C.** (see Saiklai, George C.)—763
- Schulman, J. H.** (see von Hippel, A.)—215
- Schulman, James H.** Luminescence of $(\text{Zn,Be})_2\text{SiO}_4\text{:Mn}$ and other manganese-activated phosphors—902
- Schwardtfefer, H.** Erratum: On the representation of rigid rotations—109
- Sears, W. C.** Book review—536
- Sherwood, Thomas K.** (see Brown, Gordon P.)—802
- Smith, A. E.** (see Cravath, A. M.)—309

- Smith, Alva W., J. H. Gregory, and J. T. Lynn.** Electrical resistance of iron wires and permalloy strips at radio frequencies--33
- Smoluchowski, R., C. M. Lucht, and J. M. Hurd.** Composition and diffraction effects in x-ray microradiographs--864
- Somerville, J. M.** (see Makinson, R. E. B.)--567
- Spencer, Domina Eberle** (see Moon, Parry)--506
- Spencer, R. S. and R. F. Boyer.** Thermal expansion and second-order transition effects in high polymers. III. Time effects--398
- Stockman, Harry.** Calculation of the output from non-linear mixers--110
- Stoker, R. L.** Method of determining the size of droplets dispersed in a gas--243
- Sturkey, L.** (see Heidenreich, R. D.)--127
- Szklal, George C. and Alfred C. Schroeder.** Electronic spectroscopy--763
- Tai, C. T.** Shunt and series sections of transmission line for impedance matching--44
- Teal, G. K., J. R. Fisher, and A. W. Treptow.** New bridge photo-cell employing a photo-conductive effect in silicon. Some properties of high purity silicon 879
- Thonemann, P.** (see Makinson, R. E. B.) 567
- Tobolsky, A. V.** (see Andrews, R. D.) -352
- Treptow, A. W.** (see Teal, G. K.) -879
- Tyndall, E. P. T.** (see Read, Thomas A.) -713
- Valesek, Joseph.** Book review -1130
- Van Vleck, J. H. and David Middleton.** Theoretical comparison of the visual, aural, and meter reception of pulsed signal in the presence of noise--940
- Verdery, R. B.** (see Busse, W. F.)--376
- Vinograd, J. R.** (see Cravath, A. M.)--309
- Von Karman, Theodore.** Atomic engineering--2
- Vonnegut, B. and J. L. Glathart.** Effect of temperatures on the strength and fatigue of glass rods--1082
- Wack, Paul E.** (see Guth, Eugene)--347
- Wakeham, Helmut and Edith Honold.** Hysteresis and related elastic properties of tire cords--698
- Walker, L. R. and Nelson Wax.** Non-uniform transmission lines and reflection coefficients--1043
- Waller, R. C. and W. E. Roseveare.** Fatigue failure of rayon tire cord--482
- Walsh, Thomas J.** Book review--409
- Walter, John E. and W. D. Herahberger.** Absorption of microwave by gases. II--814
- Washburn, H. W.** (see Berry, Clifford E.) -262
- Watson, John H. L.** Filmless sample mounting for the electron microscope--121
- and **K. Kaufmann.** Electron microscope examination of the microphysical properties of the polymer cuprene--996
- Wax, Nelson.** (see Walker, L. R.)--1043
- Weyl, W. A.** Use of color and fluorescence indicators for determining the structure of glass--628
- Wigner, E. P.** Theoretical physics in the metallurgical laboratory of Chicago--857; Addendum--1130
- Wilcox, Doyle E.** (see Berry, Clifford E.)--262
- Wildschut, A. J.** On the proportion of crystalline and amorphous components in stretched vulcanized rubber - 51
- Williams, Robley C.** Book review--538
- and **Ralph W. G. Wyckoff.** Applications of metallic shadow-casting to microscopy - 23
- Williams, Van Zandt** (see Barnes, R. Bowling) 532
- Wilson, J. N.** (see Cravath, A. M.)--309
- Wood, Lawrence A. and Norman Bekkedahl.** Crystallization of unvulcanized rubber at different temperatures--362
- Woodford, D. E.** (see Field, J. E.)--386
- Woods, H. L.** (see Heidenreich, R. D.) -127
- Wright, D. K., Jr.** Book reviews--312, 535
- Wyckoff, Ralph W. G.** (see Williams, Robley C.)--23
- Young, J. L., III** (see Beams, J. W.) -886
- Zener, Clarence** (see Hollomon, J. H.)--82
- and **J. H. Hollomon.** Problems in non-elastic deformation of metals--69
- Ziegler, George E.** Midwest Research Institute -a new scientific research institution--669

Subject Index to Volume 17

References with (L) are to Letters to the Editor.

- Absorption of microwaves by gases, W. D. Hershberger—495
 — Part II, John E. Walter and W. D. Hershberger - 814
 Abstracts of papers presented at meeting of Division of High Polymer Physics—405
 Accuracy of lumping in an electric circuit representing heat flow in cylindrical and spherical bodies, Victor Paschkis and Michael P. Heisler—246
 Acoustic measurements, Resonant circuit modulator for broad band acoustic measurements, Gordon Ferrie Hull, Jr.—1066
 Aeroplane propeller, Lightning recorded by, J. M. Bryant and M. Newman—555
 Alloys, Investigation of secondary phases in alloys by electron diffraction and the electron microscope, R. D. Heidenreich, L. Sturkey, and H. L. Woods—127
 Amplifiers, Design of broad band I.F. amplifiers, Richard F. Baum—519
 — Part II 721
 Amplifiers, On maximum gain-band width product in amplifiers, W. W. Hansen—109 (erratum)
 Amplifiers, Oscillation conditions in single tuned amplifiers, William R. Faust and Hugo M. Beck—749
 Analyses of mixtures of light gases by infra-red absorption, Norman D. Coggeshall and Eleanor L. Saier—450
 Analysis of multicomponent mixtures of hydrocarbons in the liquid phase by means of infra-red absorption spectroscopy, D. L. Fry, R. E. Nusbaum, and H. M. Randall—150
 Antenna, Thin cylindrical antenna: a comparison of theories, David Middleton and Ronold King -273
 Antimony, Evaporation of, Louis Harris—757 (L)
 Applications of infra-red methods in the structural examination of synthetic rubber, J. E. Field, D. E. Woodford, and S. D. Gehman—386
 Applications of metallic shadow-casting to microscopy, Robley C. Williams and Ralph W. G. Wyckoff—23
 Approximate calculation of the range of a winged jet projectile, M. Zbigniew Krzywoblocki—515
 Approximations to Planckian distributions, Parry Moon and Domina Eberle Spencer—506
 Asymptotic solutions for the normal modes in the theory of microwave propagation, C. L. Pekeris—1108
 Atomic engineering, Theodore Von Karman—2
 Atomic scientists, Federation of—108
 Austenite-Martensite transformation in steel, Electrical observations of the, Andrew W. McReynolds—823
 Bateman, Note on a paper of H. Bateman, F. W. J. Olver—1127 (L)
 Betatron, Methods for betatron or synchrotron beam removal, Eugene C. Crittenden, Jr., and William E. Parkins—444
 Bibliography of electron microscopy, Supplementary, Mary E. Rathbun, Mary J. Eastwood, and Orlan M. Arnold—759
 Body waves in water, Spark shadowgraphic study of, J. Howard McMillen and E. Newton Harvey—541
 Calculated frequency spectrum of the shot noise from a photo-multiplier tube, R. D. Sard—768
 Calculation of the output from non-linear mixers, Harry Stockman—110
 Capillary tube, Liquid rise in a, Wesley E. Brittin—37
 Cathode-ray tube for viewing continuous patterns, J. B. Johnson—891
 Cathodes, oxide coated, Dissociation energies of surface films of various oxides as determined by emission measurements of, Harold Jacobs—596
 Centimeter waves, New method for measuring dielectric constant and loss in the range of, S. Roberts and A. von Hippel—610
 Centrifugal fields, Production of high centrifugal fields, J. W. Beams, J. L. Young, III, and J. W. Moore—886
 Circuit modulator, Resonant circuit modulator for broad band acoustic measurements, Gordon Ferris Hull, Jr. - 1066
 Color and fluorescence indicators for determining the structure of glass, Use of, W. A. Weyl—628
 Composition and diffraction effects in x-ray microradiographs, R. Smoluchowski, C. M. Lucht, and J. M. Hurd—864
 Computer for solving linear simultaneous equations, Clifford E. Berry, Doyle E. Wilcox, Sibyl M. Rock, and H. W. Washburn—262
 Computer for solving linear simultaneous equations, C. C. Eaglesfield; Clifford E. Berry—1125 (L).
 Conductors, Radar reflections from long conductors, F. Bloch, M. Hamermesh, and M. Phillips—1015
 Continuous patterns, Cathode-ray tube for viewing continuous patterns, J. B. Johnson—891
 Coordination number in solid solutions—a mathematical paradox, Douglas Rennie Hudson—1126 (L)
 Cork, Elastic properties of. I. Stress relaxation of compressed cork, S. L. Dart and Eugene Guth—314
 Correction of diffraction amplitudes for Lorentz and polarization factors, M. J. Buerger and Gilbert E. Klein—285
 Correction of diffraction amplitudes for Lorentz and polarization factors, M. J. Buerger and Gilbert E. Klein—743 (erratum)
 Creep, Plastic flow, creep, and stress relaxation. Part I—Plastic flow, Charles Mack—1086
 — Part II—Creep—1093
 — Part III—Creep and elastic after-effect—1101
 Crystal structures, Machine for the application of sand in making Fourier projections of crystal structures, Dan McLachlan, Jr., and E. F. Champayne—1006
 Crystalline films, x-ray method for measuring the thickness of thin crystalline films, A. Eisenstein—874
 Crystallization of unvulcanized rubber at different temperatures, Lawrence A. Wood and Norman Bekkedahl—362

- Cuprene, Electron microscope examination of the microphysical properties of the polymer cuprene, John H. L. Watson and K. Kaufmann—996
- Cylinders, Induction heating of hollow metallic cylinders, Andrew Gemant—195
- Cylindrical dipoles, Measured impedance of, D. D. King—844
- Debunching, Elementary treatment of longitudinal debunching in a velocity modulation system, Eugene Feenberg—852
- Deformation of metals, Problems in non-elastic deformation of metals, Clarence Zener and J. H. Hollomon—69
- Depreciation of fluorescent lamps without passage of current, H. C. Froelich—573
- Design of broad band I. F. amplifiers, Richard F. Baum—519
- Part II—721
- Determination of the temperature rise and the maximum safe current through multiconductor electric cables, Haig P. Iskenderian and William J. Horvath—255
- Dielectric constant and loss in the range of centimeter waves, New method for measuring, S. Roberts and A. von Hippel—610
- Dielectric loss measuring technique, Microwave, William R. MacLean—558
- Dielectrics, Formation of ionized water films on dielectrics under conditions of high humidity, Robert F. Field—318
- Dimethylsiloxane polymers, Viscometric investigation of, Arthur J. Barry—1020
- Dissociation energies of surface films of various oxides as determined by emission measurements of oxide coated cathodes, Harold Jacobs—596
- Division of High Polymer Physics, Abstracts—405
- Dynamic characteristics of rubber supports from vibration table data, Ernest G. Chilton—492
- Effect of hydrostatic pressure on plastic flow under shearing stress, P. W. Bridgman—692
- Effect of temperature on the strength and fatigue of glass rods, B. Vonnegut and J. L. Glathart—1082
- Effect of water on the strength of glass, T. C. Baker and F. W. Preston—179
- Elastic limit under dynamic loading, V. Bargmann and H. F. Ludloff—63 (L)
- Elastic properties of cork. I. Stress relaxation of compressed cork, S. L. Dart and Eugene Guth—314
- Elastomers, Tackiness of GR-S and other elastomers, W. F. Busse, J. M. Lambert, and R. B. Verdery—376
- Electric cables, Determination of the temperature rise and the maximum safe current through multiconductor electric cables, Haig P. Iskenderian and William J. Horvath—255
- Electric circuit, Accuracy of lumping in an electric circuit representing heat flow in cylindrical and spherical bodies, Victor Paschkis and Michael P. Heisler—246
- Electrical observations of the Austenite-Martensite transformation in steel, Andrew W. McReynolds—823
- Electrical resistance of iron wires and permalloy strips at radiofrequencies, Alva W. Smith, J. H. Gregory, and J. T. Lynn—33
- Electron beam, Interchange of energy between an electron beam and an oscillating electric field, Jess Marcum—4
- Electron beam, Theory of small signal bunching in a parallel electron beam of rectangular cross section, Eugene Feenberg and David Feldman—1025
- Electron diffraction cameras, Improvement of resolution in, James Hillier and R. F. Baker—12
- Electron diffraction and the electron microscope, Investigation of secondary phases in alloys by, R. D. Heidenreich, L. Sturkey, and H. L. Woods—127
- Electron microscope examination of the microphysical properties of the polymer cuprene, John H. L. Watson and K. Kaufmann—996
- Electron microscope, Filmless sample mounting for the, John H. L. Watson—121
- Electron microscope, Further improvement in the resolving power of the, James Hillier—307
- Electron microscope, Method for preparing rubber latex specimens for the, R. H. Kelsey and E. E. Hanson—675
- Electron microscope, Preparing pigments for the, A. M. Cravath—1125 (L)
- Electron Microscope Society of America, Report of the Committee on Resolution of the—989
- Electron Microscope Society of America. Third Annual meeting—66
- Electron microscope specimens, Preparation of, for determination of particle size distribution in aqueous suspensions, A. M. Cravath, A. E. Smith, J. R. Vinograd, and J. N. Wilson—309
- Electron microscope, Study of distortion in electron microscope projection lenses, James Hillier—411
- Electron microscopy, Supplementary bibliography of, Mary E. Rathbun, Mary J. Eastwood, and Orlan M. Arnold—759
- Electronic spectroscopy, George C. Sziklai and Alfred C. Schroeder—763
- Electrostatic and tensile properties of rubber and GR-S at elevated temperatures, R. S. Havenhill, H. C. O'Brien, and J. J. Rankin—338
- Elementary treatment of longitudinal debunching in a velocity modulation system, Eugene Feenberg—852
- Emission measurements of oxide coated cathodes, Dissociation energies of surface films of various oxides as determined by, Harold Jacobs—596
- Energy build-up in magnetrons, Lloyd P. Hunter—833
- Equation of state, Significance of the equation of state for rubber, Eugene Guth, Paul E. Wack, and Robert L. Anthony—347
- Evaporation of antimony, Louis Harris—757 (L)
- Experimental investigation of forced vibrations in a mechanical system having a non-linear restoring force, Carl A. Ludeke—603
- Fatigue failure of rayon tire cord, R. C. Waller and W. E. Roseveare—482
- Fatigue of glass under static loads, T. C. Baker and F. W. Preston—170
- Fatigue modulus of glass, J. L. Glathart and F. W. Preston—189
- Federation of atomic scientists—108

- Filmless sample mounting for the electron microscope, John H. L. Watson -121
- Film-lubrication between spherical surfaces: with an application to the theory of the four-ball lubricant testing instrument, J. Howlett—137
- Fine structure of viscose rayon, H. G. Ingersoll—924
- Flow of gases in pipes at low pressures, Gordon P. Brown, Albert DiNardo, George K. Cheng, and Thomas K. Sherwood - 802
- Flow in solids, General relations of flow in solids and their application to the plastic behavior of tire cords, W. James Lyons—472
- Fluorescent lamps, Depreciation of, without passage of current, H. C. Froelich -573
- Formation of ionized water films on dielectrics under conditions of high humidity, Robert F. Field—318
- Four-ball lubricant testing instrument, Film-lubrication between spherical surfaces: with an application to the theory of, J. Howlett—137
- Fourier projections, Machine for the application of sand in making Fourier projections of crystal structures, Dan McLachlan, Jr., and E. F. Champayne -1006
- FP-54 electrometer tube, Improvements in the stability of the, J. M. Lafferty and K. H. Kingdon—894
- Frequency discriminator, George G. Bruck—213 (L)
- Frequency spectrum, Calculated frequency spectrum of the shot noise from a photo-multiplier tube, R. D. Sard - 768
- Further improvement in the resolving power of the electron microscope, James Hillier—307
- Gas discharge tubes, Magnetically-controlled, R. E. B. Makinson, J. M. Somerville, K. Rachel Makinson, and P. Thonemann -567
- Gas, Method of determining the size of droplets dispersed in a, R. L. Stoker—243
- Gases, Absorption of microwaves by gases. II, John E. Walter and W. D. Hershberger—814
- Gases, Flow of gases in pipes at low pressures, Gordon P. Brown, Albert DiNardo, George K. Cheng, and Thomas K. Sherwood—802
- Gases, Permeability of different rubbers to gases and its relation to diffusivity and solubility, G. J. van Amerongen—972
- General relations of flow in solids and their application to the plastic behavior of tire cords, W. James Lyons—472
- Glass, Effect of water on the strength of, T. C. Baker and F. W. Preston—179
- Glass, Fatigue of glass under static loads, T. C. Baker and F. W. Preston—170
- Glass, Fatigue modulus of, J. L. Glathart and F. W. Preston—189
- Glass-metal seals, Stresses in cylindrical glass-metal seals glass inside, Albert W. Hull—685
- Glass rods, Effect of temperature on the strength and fatigue of, B. Vonnegut and J. L. Glathart—1082
- Glass rods, Wide range static strength testing apparatus for, T. C. Baker and F. W. Preston—162
- Glass, Use of color and fluorescence indicators for determining the structure of, W. A. Weyl—628
- Graphical method for determining particle trajectories, William E. Parkins and Eugene C. Crittenden, Jr.—447
- GR-S, Electrostatic and tensile properties of rubber and GR-S at elevated temperatures, R. S. Havenhill, H. C. O'Brien, and J. J. Rankin—338
- GR-S, Tackiness of, and other elastomers, W. F. Busse, J. M. Lambert, and R. B. Verdery—376
- Harker syntheses, Interpretation of, M. J. Buerger—579
- Heat flow, Accuracy of lumping in an electric circuit representing heat flow in cylindrical and spherical bodies, Victor Paschkis and Michael P. Heisler—246
- Human aorta, Pressure-volume relation for cylindrical tubes with elastomeric walls: the human aorta, Allen L. King—501
- Hydrocarbons, Analysis of multicomponent mixtures of hydrocarbons in the liquid phase by means of infra-red absorption spectroscopy, D. L. Fry, R. E. Nusbaum, and H. M. Randall—150
- Hysteresis and related elastic properties of tire cords, Helmut Wakeham and Edith Honold—698
- Impedance matching, Shunt and series sections of transmission line for, C. T. Tai—44
- Impedance, Measured impedance of cylindrical dipoles, D. D. King - 844
- Improvement of resolution in electron diffraction cameras, James Hillier and R. F. Baker—12
- Improvements in the stability of the FP-54 electrometer tube, J. M. Lafferty and K. H. Kingdon—894
- Induction heating of hollow metallic cylinders, Andrew Gemant—195
- Infra-red absorption, Analyses of mixtures of light gases by, Norman D. Coggeshall and Eleanor L. Saier—450
- Infra-red absorption spectroscopy, Analysis of multicomponent mixtures of hydrocarbons in the liquid phase by means of, D. L. Fry, R. E. Nusbaum, and H. M. Randall—150
- Infra-red, Applications of infra-red methods in the structural examination of synthetic rubber, J. E. Field, D. E. Woodford, and S. D. Gehman—386
- Infra-red spectroscopy, Small prism, R. Bowling Barnes, Robert S. McDonald, and Van Zandt Williams—532 (erratum)
- Integral equations of potential theory, Some, H. Bateman—91
- Interchange of energy between an electron beam and an oscillating electric field, Jess Marcum—4
- Internal friction and plastic extension of zinc single crystals, Thomas A. Read and E. P. T. Tyndall—713
- Interpolation with the aid of a plot of first differences, Gordon S. Fulcher—617
- Interpretation of Harker syntheses, M. J. Buerger—579
- Investigation, Experimental, of forced vibrations in a mechanical system having a non-linear restoring force, Carl A. Ludeke—603
- Investigation of secondary phases in alloys by electron diffraction and the electron microscope, R. D. Heidenreich, L. Sturkey, and H. L. Woods—127
- Ionized water films, Formation of, on dielectrics under conditions of high humidity, Robert F. Field—318

- Jet projectile, Approximate calculation of the range of a winged jet projectile, M. Zbigniew Krzywoblocki—515
- Journal of Applied Physics, Reconversion of the—1
- Laplace's equation, Numerical solution of, in composite rectangular areas, Max M. Frocht—730
- Lead loss, Methods of minimizing lead loss in emissivity and resistivity determinations, L. Malter and D. B. Langmuir—456
- Light gases, Analyses of mixtures of light gases by infrared absorption, Norman D. Coggeshall and Eleanor L. Saier—450
- Light scattering measurements, Photoelectric instrument for light scattering measurements and a differential refractometer, P. P. Debye—392
- Lightning recorded by aeroplane propeller, J. M. Bryant and M. Newman—555
- Linear simultaneous equations, Computer for solving, Clifford E. Berry, Doyle E. Wilcox, Sibyl M. Rock, and H. W. Washburn—262
- Linear simultaneous equations, Computer for solving, C. C. Eaglesfield; Clifford E. Berry—1125 (L)
- Linear systems, Transient analysis of, using underwater explosion waves, M. F. M. Osborne and J. L. Carter—871
- Liquid rise in a capillary tube, Wesley E. Brittin—37
- Lorentz and polarization factors, Correction of diffraction amplitudes for, M. J. Buerger and Gilbert E. Klein—285—743 (erratum)
- Luminescence of $(\text{Zn, Be})_2\text{SiO}_4\text{:Mn}$ and other manganese-activated phosphors, James H. Schulman—902
- Machine for the application of sand in making Fourier projections of crystal structures, Dan McLachlan, Jr., and E. F. Champayne—1006
- Magnetically-controlled gas discharge tubes, R. E. B. Makinson, J. M. Somerville, K. Rachel Makinson, and P. Thonemann—567
- Magnetrons, Energy build-up in, Lloyd P. Hunter—833
- Maximum gain-band width product in amplifiers, W. W. Hansen—109 (erratum)
- Measured impedance of cylindrical dipoles, D. D. King—844
- Mechanical system having a non-linear restoring force, Experimental investigation of forced vibrations in a, Carl A. Ludeke—603
- Metallic shadow-casting, Applications of metallic shadow-casting to microscopy, Robley C. Williams and Ralph W. G. Wyckoff—23
- Metallurgical laboratory of Chicago, Theoretical physics in the, E. P. Wigner—857; Addendum—1130
- Metals, Problems in fracture of, J. H. Hollomon and C. Zener—82
- Metals, Problems in non-elastic deformation of, Clarence Zener and J. H. Hollomon—69
- Method of determining the size of droplets dispersed in a gas, R. L. Stoker—243
- Method for preparing rubber latex specimens for the electron microscope, R. H. Kelsey and E. E. Hanson—675
- Methods for betatron or synchrotron beam removal, Eugene C. Crittenden, Jr., and William E. Parkins—444
- Methods of minimizing lead loss in emissivity and resistivity determinations, L. Malter and D. B. Langmuir—456
- Microscopy, Applications of metallic shadow-casting to, Robley C. Williams and Ralph W. G. Wyckoff—23
- Microwave dielectric loss measuring technique, William R. MacLean—558
- Microwave propagation, Asymptotic solutions for the normal modes in the theory of, C. L. Pekeris—1108
- Microwaves, Absorption of microwaves by gases, W. D. Herschberger—495
- Part II, John E. Walter and W. D. Herschberger—814
- Microwaves, Perturbation theory of the normal modes for an exponential M -curve in non-standard propagation of microwaves, C. L. Pekeris—678
- Midwest Research Institute a new scientific research institution, George E. Ziegler—669
- Millimeter-wave reflex oscillator, J. M. Lafferty—1061
- New bridge photo-cell employing a photo-conductive effect in silicon. Some properties of high purity silicon, G. K. Teal, J. R. Fisher, and A. W. Treptow—879
- New electrolytic selenium photo-cell, A. von Hippel, J. H. Schulman, and E. S. Rittner—215
- New method for measuring dielectric constant and loss in the range of centimeter waves, S. Roberts and A. von Hippel—610
- Noise, Response of biased, saturated linear and quadratic rectifiers to random noise, David Middleton—778
- Non-linear mixers, Calculation of the output from, Harry Stockman—110
- Non-uniform transmission lines and reflection coefficients, L. R. Walker and N. Wax—1043
- Noised-in cavity, Resonant frequencies of the, Ernest Mayer—1046
- Note on a paper of H. Bateman, F. W. J. Olver—1127 (L)
- Novel form of refrigerator, Jenny E. Rosenthal—62 (L)
- Numerical solution of Laplace's equation in composite rectangular areas, Max M. Frocht—730
- Oils, Thermal polymerization of drying oils, H. E. Adams and P. O. Powers—325
- Oscillation conditions in single tuned amplifiers, William R. Faust and Hugo M. Beck—749
- Oscillator, Millimeter-wave reflex oscillator, J. M. Lafferty—1061
- Oxide cathodes, Pulsed properties of, Edward A. Coomes—647
- Oxide cathodes, Study of, by x-ray diffraction methods. Part I. Methods, conversion studies, and thermal expansion coefficients, A. Eisenstein—434
- Part II. Oxide coating composition—654
- Oxide coated cathode literature, 1940-1945, John P. Blewett—643
- Oxide coated cathodes, Dissociation energies of surface films of various oxides as determined by emission measurements of, Harold Jacobs—596
- Oxide coated cathodes, Studies of the interface of, A. Fineman and A. Eisenstein—663
- Ozone, Recent ozone investigations, Arthur W. Ewell—908

- Parafocusing properties of microcrystalline powder layers in x-ray diffraction applied to the design of x-ray goniometers, J. C. M. Brentano—420
- Parallel plate plastometer, Theory and application of the, G. J. Dienes and H. F. Klemm—458
- Particle accelerator, Study of a certain type of resonant cavity and its application to a charged particle accelerator, E. S. Akeley—1056
- Particle size determination from x-ray line broadening, L. S. Birks and H. Friedman—687
- Particle size distribution, Preparation of electron microscope specimens for determination of particle size distribution in aqueous suspensions, A. M. Cravath, A. E. Smith, J. R. Vinograd, and J. N. Wilson—309
- Particle trajectories, Graphical method for determining particle trajectories, William E. Parkins and Eugene C. Crittenden, Jr.—447
- Permeability of different rubbers to gases and its relation to diffusivity and solubility, G. J. van Amerongen—972
- Perturbation theory of the normal modes for an exponential M -curve in non-standard propagation of microwaves, C. L. Pekeris—678
- Phosphors, Luminescence of $(Zn, Be)_2SiO_4:Mn$ and other manganese-activated phosphors, James H. Schulman—902
- Phosphors, Thermoluminescence and conductivity of, Robert C. Herman and Charles F. Meyer—743
- Photo-cell, New bridge photo-cell employing a photo-conductive effect in silicon. Some properties of high purity silicon, G. K. Teal, J. R. Fisher, and A. W. Treptow—879
- Photo-cell, New electrolytic selenium photo-cell, A. von Hippel, J. H. Schulman, and E. S. Rittner—215
- Photoelectric instrument for light scattering measurements and a differential refractometer, P. P. Debye—392
- Photo-multiplier tube, Calculated frequency spectrum of the shot noise from a, R. D. Sard—768
- Picturesque oxidation of molybdenum vacuum-tube leads, Herman A. Liebhafsky—901
- Pigments, Preparing pigments for the electron microscope, A. M. Cravath—1125 (L)
- Pipes, Transient temperatures around heating pipes maintained at constant temperature, Andrew Gemant—1076
- Planckian distributions, Approximations to, Parry Moon and Domina Eberle Spencer—506
- Plane rotations in n dimensions, Peter T. Landsberg—60 (L)
- Plastic extension, Internal friction and plastic extension of zinc single crystals, Thomas A. Read and E. P. T. Tyndall—713
- Plastic flow, creep, and stress relaxation. Part I—Plastic flow, Charles Mack—1086
- Part II—Creep—1093
- Part III—Creep and elastic after-effect—1101
- Plastic flow, Effect of hydrostatic pressure on plastic flow under shearing stress, P. W. Bridgman—692
- Plastic flow of steel, Studies of, especially in two-dimensional compression, P. W. Bridgman—225
- Polymer cuprene, Electron microscope examination of the microphysical properties of the, John H. L. Watson and K. Kaufmann—996
- Polymer, Division of High Polymer Physics, Abstracts—405
- Polymerization, Thermal polymerization of drying oils, H. E. Adams and P. O. Powers—325
- Polymers, Thermal expansion and second-order transition effects in high polymers. III. Time effects, R. S. Spencer and R. F. Boyer—398
- Polymers, Viscometric investigation of dimethylsiloxane polymers, Arthur J. Barry—1020
- Power function, Fidel Alsina Fuertes—712 (L)
- Preparation of electron microscope specimens for determination of particle size distribution in aqueous suspensions, A. M. Cravath, A. E. Smith, J. R. Vinograd, and J. N. Wilson—309
- Preparing pigments for the electron microscope, A. M. Cravath—1125 (L)
- Pressure-volume relation for cylindrical tubes with elastomeric walls: the human aorta, Allen L. King—501
- Problems in fracture of metals, J. H. Hollomon and C. Zener—82
- Problems in non-elastic deformation of metals, Clarence Zener and J. H. Hollomon—69
- Production of high centrifugal fields, J. W. Beams, J. L. Young, III, and J. W. Moore—886
- Properties of welded contact germanium rectifiers, H. Q. North—912
- Proportion of crystalline and amorphous components in stretched vulcanized rubber, A. J. Wildschut—51
- Pulsed properties of oxide cathodes, Edward A. Coomes—647
- Pulsed signal, Theoretical comparison of the visual, aural, and meter reception of pulsed signal in the presence of noise, J. H. Van Vleck and David Middleton—940
- Pyrex glass, Secondary electron emission of, C. W. Mueller—62 (erratum)
- Radar reflections from long conductors, F. Bloch, M. Hamermesh, and M. Phillips—1015
- Radiofrequencies, Electrical resistance of iron wires and permalloy strips at, Alva W. Smith, J. H. Gregory, and J. T. Lynn—33
- Range of centimeter waves, New method for measuring dielectric constant and loss in the, S. Roberts and A. von Hippel—610
- Rayon, Fine structure of viscose rayon, H. G. Ingersoll—924
- Rayon tire cord, Fatigue failure of, R. C. Waller and W. E. Roseveare—482
- Recent ozone investigations, Arthur W. Ewell—908
- Reconversion of the Journal of Applied Physics—1
- Rectifiers, Properties of welded contact germanium rectifiers, H. Q. North—912
- Rectifiers, Response of biased, saturated linear and quadratic rectifiers to random noise, David Middleton—778
- Reflection coefficients, Non-uniform transmission lines and, L. R. Walker and N. Wax—1043

- Refractometer, Photoelectric instrument for light scattering measurements and a differential refractometer, P. Debye—392
- Refrigerator, Novel form of, Jenney E. Rosenthal—62 (L)
- Relation between nodal positions and standing wave ratio in a composite transmission system, Eugene Feenberg—530
- Report of the Committee on Resolution of the Electron Microscope Society of America—989
- Representation of rigid rotations, H. Schwerdtfeger—109 (erratum)
- Resonant cavity, Study of a certain type of, and its application to a charged particle accelerator, E. S. Akeley—1056
- Resonant circuit modulator for broad band acoustic measurements, Gordon Ferrie Hull, Jr.—1066
- Resonant frequencies of the nosed-in cavity, Ernest Mayer—1046
- Response of biased, saturated linear and quadratic rectifiers to random noise, David Middleton—778
- Rigid rotations, On the representation of, H. Schwerdtfeger—109 (erratum)
- Rubber, Applications of infra-red methods in the structural examination of synthetic rubber, J. E. Field, D. E. Woodford, and S. D. Gehman—386
- Rubber, Crystallization of unvulcanized rubber at different temperatures, Lawrence A. Wood and Norman Bekkedahl—362
- Rubber, Electrostatic and tensile properties of rubber and GR-S at elevated temperatures, R. S. Havenhill, H. C. O'Brien, and J. J. Rankin—338
- Rubber latex, Method for preparing rubber latex specimens for the electron microscope, R. H. Kelsey and E. E. Hanson—675
- Rubber, On the proportion of crystalline and amorphous components in stretched vulcanized rubber, A. J. Wildschut—51
- Rubber, Significance of the equation of state for rubber, Eugene Guth, Paul E. Wack, and Robert L. Anthony—347
- Rubber supports, Dynamic characteristics of rubber supports from vibration table data, Ernest G. Chilton—492
- Rubber vulcanizates, Theory of permanent set at elevated temperatures in natural and synthetic rubber vulcanizates, R. D. Andrews, A. V. Tobolsky, and E. E. Hanson—352
- Rubbers, Permeability of different rubbers to gases and its relation to diffusivity and solubility, G. J. van Amerongen—972
- Sand, Machine for the application of sand in making Fourier projections of crystal structures, Dan McLachlan, Jr., and E. F. Champayne—1006
- Secondary electron emission of Pyrex glass, C. W. Mueller—62 (erratum)
- Shells, Vibrations of shallow spherical shells, Eric Reissner—1038
- Shunt and series sections of transmission line for impedance matching, C. T. Tai—44
- Signal bunching, Theory of small signal bunching in a parallel electron beam of rectangular cross section, Eugene Feenberg and David Feldman—1025
- Significance of the equation of state for rubber, Eugene Guth, Paul E. Wack, and Robert L. Anthony—347
- Silicon, New bridge photo-cell employing a photo-conductive effect in silicon. Some properties of high purity silicon, G. K. Teal, J. R. Fisher, and A. W. Treptow—879
- Small prism infra-red spectroscopy, R. Bowling Barnes, Robert S. McDonald, and Van Zandt Williams—532 (erratum)
- Solid solutions, Coordination number in solid solutions—a mathematical paradox, Douglas Rennie Hudson—1126 (L)
- Some integral equations of potential theory, H. Bateman—91
- Spark shadowgraphic study of body waves in water, J. Howard McMillen and E. Newton Harvey—541
- Spectroscopy, Electronic, George C. Sziklai and Alfred C. Schroeder—763
- Static strength testing apparatus, Wide range, for glass rods, T. C. Baker and F. W. Preston—162
- Steel, Electrical observations of the Austenite-Martensite transformation in, Andrew W. McReynolds—823
- Steel, Studies of plastic flow of steel, especially in two-dimensional compression, P. W. Bridgman—225
- Steels, Tensile properties of several special steels and certain other materials under pressure, P. W. Bridgman—201
- Stress relaxation, Elastic properties of cork. I. Stress relaxation of compressed cork, S. L. Dart and Eugene Guth—314
- Stress relaxation, Plastic flow, creep, and stress relaxation. Part I—Plastic flow, Charles Mack—1086
— Part II—Creep—1093
— Part III—Creep and elastic after-effect—1101
- Stresses in cylindrical glass-metal seals with glass inside, Albert W. Hull—685
- Structure of glass, Use of color and fluorescence indicators for determining the, W. A. Weyl—628
- Studies of the interface of oxide coated cathodes, A. Fine-ma, and A. Eisenstein—663
- Studies of plastic flow of steel, especially in two-dimensional compression, P. W. Bridgman—225
- Study of a certain type of resonant cavity and its application to a charged particle accelerator, E. S. Akeley—1056
- Study of distortion in electron microscope projection lenses, James Hillier—411
- Study of oxide cathodes by x-ray diffraction methods. Part I. Methods, conversion studies, and thermal expansion coefficients, A. Eisenstein—434
— Part II. Oxide coating composition—654
- Supplementary bibliography of electron microscopy, Mary E. Rathbun, Mary J. Eastwood, and Orlan M. Arnold—759
- Surface films of various oxides as determined by emission measurements of oxide coated cathodes, Dissociation energies of, Harold Jacobs—596
- Surface tension, Gerald Pickett—61 (L)

- Synchrotron, Methods for betatron or synchrotron beam removal, Eugene C. Crittenden, Jr., and William E. Parkins—444
- Tackiness of GR-S and other elastomers, W. F. Busse, J. M. Lambert, and R. B. Verdery—376
- Temperatures, Transient temperatures around heating pipes maintained at constant temperature, Andrew Gemant—1076
- Tensile properties of several special steels and certain other materials under pressure, P. W. Bridgman—201
- Theoretical comparison of the visual, aural, and meter reception of pulsed signal in the presence of noise, J. H. Van Vleck and David Middleton—940
- Theoretical physics in the metallurgical laboratory of Chicago, E. P. Wigner—857; Addendum—1130
- Theory and application of the parallel plate plastometer, G. J. Dienes and H. F. Klemm—458
- Theory of permanent set at elevated temperatures in natural and synthetic rubber vulcanizates, R. D. Andrews, A. V. Tobolsky, and E. E. Hanson—352
- Theory of small signal bunching in a parallel electron beam of rectangular cross section, Eugene Feenberg and David Feldman—1025
- Thermal expansion and second-order transition effects in high polymers. III. Time effects, R. S. Spencer and R. F. Boyer—398
- Thermal polymerization of drying oils, H. E. Adams and P. O. Powers—325
- Thermoluminescence and conductivity of phosphors, Robert C. Herman and Charles F. Meyer—743
- Thin cylindrical antenna: a comparison of theories, David Middleton and Ronald King—273
- Tire cords, General relations of flow in solids and their application to the plastic behavior of, W. James Lyons—472
- Tire cords, Hysteresis and related elastic properties of, Helmut Wakeham and Edith Honold—698
- Transient analysis of linear systems, using underwater explosion waves, M. F. M. Osborne and J. L. Carter—871
- Transient temperatures around heating pipes maintained at constant temperature, Andrew Gemant—1076
- Transmission lines, Non-uniform, and reflection coefficients, L. R. Walker and N. Wax—1043
- Transmission system, Relation between nodal positions and standing wave ratio in a composite transmission system, Eugene Feenberg—530
- Tubes, Magnetically-controlled gas discharge tubes, R. E. B. Makinson, J. M. Somerville, K. Rachel Makinson, and P. Thonemann—567
- Underwater explosion waves, Transient analysis of linear systems, using underwater explosion waves, M. F. M. Osborne and J. L. Carter—871
- Use of color and fluorescence indicators for determining the structure of glass, W. A. Weyl—628
- Vacuum-tube leads, Picturesque oxidation of molybdenum vacuum-tube leads, Herman A. Liebhafsky—901
- Velocity modulation system, Elementary treatment of longitudinal debunching in a, Eugene Feenberg—852
- Vibrations, forced, in a mechanical system having a non-linear restoring force, Experimental investigation of, Carl A. Ludeke—603
- Vibrations of shallow spherical shells, Eric Reissner—1038
- Viscometric investigation of dimethylsiloxane polymers, Arthur J. Barry—1020
- Viscose rayon, Fine structure of, H. G. Ingersoll—924
- Water, Effect of water on the strength of glass, T. C. Baker and F. W. Preston—179
- Wide range static strength testing apparatus for glass rods, T. C. Baker and F. W. Preston—162
- Wires, Electrical resistance of iron wires and permalloy strips at radiofrequencies, Alva W. Smith, J. H. Gregory, and J. T. Lynn—33
- X-ray diffraction methods, Study of oxide cathodes by x-ray diffraction methods. Part I. Methods, conversion studies, and thermal expansion coefficients, A. Eisenstein—434
- Part II. Oxide coating composition—654
- X-ray diffraction, Parafofocusing properties of microcrystalline powder layers in x-ray diffraction applied to the design of x-ray goniometers, J. C. M. Brentano—420
- X-ray goniometers, Parafofocusing properties of microcrystalline powder layers in x-ray diffraction applied to the design of x-ray goniometers, J. C. M. Brentano—420
- X-ray line broadening, Particle size determination from, L. S. Birks and H. Friedman—687
- X-ray method for measuring the thickness of thin crystalline films, A. Eisenstein—874
- X-ray microradiographs, Composition and diffraction effects in, R. Smoluchowski, C. M. Lucht, and J. M. Hurd—864
- Zinc single crystals, Internal friction and plastic extension of, Thomas A. Read and E. P. T. Tyndall—713
- (Zn, Be)₂SiO₄:Mn and other manganese-activated phosphors, Luminescence of, James H. Schulman—902

**Indian Agricultural Research
Institute Library, New Delhi**

1. Books are issued for 14 days, beyond
if for next 7 days a fine of 15 paise
& after wards 25 paise per day shall
be charged.
2. Books may be renewed at the discre-
tion of the head Library Services.
3. Dog-earing of pages of a book,
marking of writing there in with ink
or pencil, tearing or taking out its
pages or otherwise damaging it will
constitute an injury to a book.
4. Unless a borrower points out the
injury at the time of borrowing the
book, he shall be required to replace
the book or pay its price, if detected
at the time of return.

**HELP US TO KEEP THIS BOOK
FRESH AND CLEAN**

Indian Agricultural Research Institute (Pusa)
LIBRARY, NEW DELHI-110012

This book can be issued on or after _____

Return Date	Return Date